

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 23:10:30 ; Search time 23 Seconds  
(without alignments)  
1805.122 Million cell updates/sec

Title: US-10-046-433-40

Perfect score: 5506  
Sequence: 1 MAEPGSHHLSARVGRTER.....LGRSNHLPRLGLMDLTQCR 1001

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt-40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	251	4.6	1609	1	LMG1_HUMAN
2	225	4.1	1607	1	LMG1_MOUSE
3	224.5	4.1	1639	1	LMG1_DROME
4	211.5	3.8	713	1	TSN4_GIALA
5	209.5	3.8	1581	1	LMG3_MOUSE
6	209	3.8	1877	1	PKC5_MOUSE
7	208.5	3.8	1696	1	PKC5_BRACL
8	193	3.5	687	1	VS41_GIALA
9	192	3.5	1557	1	LMG3_MOUSE
10	192	3.5	1587	1	LMG3_HUMAN
11	174	3.2	1786	1	LMB1_HUMAN
12	172.5	3.1	3106	1	TS11_GIALA
13	171.5	3.1	1786	1	LMB1_MOUSE
14	170	3.1	1680	1	FUR2_DROME
15	169	3.1	2768	1	THYX_RAT
16	168.5	3.1	2768	1	LMG2_HUMAN
17	167	3.0	610	1	LMG2_HUMAN
18	166.5	3.0	3110	1	LMG1_MOUSE
19	163	3.0	3084	1	LMG1_MOUSE
20	162	2.9	1246	1	TMV2_MOUSE
21	161.5	2.9	2524	1	NOTC_XENLA
22	160.5	2.9	1790	1	LMB1_DROME
23	160.5	2.9	3672	1	LMG2_MOUSE
24	160	2.9	993	1	EPB3_MOUSE
25	160	2.9	1169	1	EPB3_MOUSE
26	159	2.9	1169	1	YK82_YEAST
27	158	2.9	1218	1	JAG1_HUMAN
28	158	2.9	2871	1	FBN1_BOVIN
29	158	2.9	3712	1	LMG2_HUMAN
30	157.5	2.9	1193	1	LMG2_HUMAN
31	157.5	2.9	2907	1	FBN2_MOUSE
32	157.5	2.9	2911	1	FBN2_HUMAN
33	157.5	2.9	3075	1	LMG1_HUMAN

34	157	2.9	4393	1	PCBM_HUMAN
35	156	2.8	1219	1	JAG1_RAT
36	155.5	2.8	1242	1	JAG1_BRARE
37	155	2.8	998	1	EPB3_HUMAN
38	155	2.8	1218	1	EPB3_MOUSE
39	154.5	2.8	4544	1	JAG1_MOUSE
40	154	2.8	2471	1	THYX_MOUSE
41	154	2.8	2766	1	THYX_MOUSE
42	154	2.8	2871	1	FBN1_MOUSE
43	152.5	2.8	974	1	EPB3_XENLA
44	151	2.7	3695	1	LMG5_HUMAN
45	151	2.7	5376	1	ZAN_MOUSE

## ALIGNMENTS

RESULT 1	LMG1_HUMAN	STANDARD	PRT	1609 AA.
AC	P11047			
DT	01-JUL-1989 (rel. 11, Created)			
DT	01-NOV-1991 (rel. 20, last sequence update)			
DT	15-JUN-2002 (rel. 41, last annotation update)			
DE	Laminin gamma-1 chain precursor (Laminin B2 chain).			
GN	LMG1 OR LAMB2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91093128; PubMed=1989895;			
RA	Kallunki T., Ikonen J., Chow L.T., Tryggvason K.;			
RT	"Structure of the human laminin B2 chain gene reveals extensive divergence from the laminin B1 chain gene."			
RT	J. Biol. Chem. 266:221-228(1991).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=86198245; PubMed=3360804;			
RX	Pikkarainen T., Kallunki T., Tryggvason K.;			
RT	"Human laminin B2 chain. Comparison of the complete amino acid sequence with the B1 chain reveals variability in sequence homology between different structural domains."			
RT	J. Biol. Chem. 263:6751-6758(1988).			
RL	[3]			
RN	SEQUENCE OF 1393-1609 FROM N.A.			
RP	MEDLINE=89169663; PubMed=3234037;			
RX	Fukushima Y., Pikkarainen T., Kallunki T., Eddy R.L., Byers M.G.;			
RA	Haley L.V., Henry W.M., Tryggvason K., Shows T.B.;			
RT	"Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of the gene to chromosome region 19q25--q31."			
RT	Cytogenet. Cell Genet. 48:137-141(1988).			
RL	[4]			
RN	SEQUENCE OF 1282-1609 FROM N.A.			
RP	TISUP-endothelial cells;			
RC	MEDLINE=92216129; PubMed=1806043;			
RX	Santos C.U.S., Sabbaga J., Brentani R.;			
RA	"Differences in human laminin B2 sequences."			
RT	DNA Seq. 1:275-277(1991).			
RL	-1- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.			
CC	-1- SUPPLEMENT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.			
CC	THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-2 (MEROSTIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSTIN), LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).			
CC	-1- SUBCELLULAR LOCATION: Extracellular.			

[illegible]

Query Match 4.6% Score 251; DB 1; Length 1609;  
 Best Local Similarity 19.6% Pred. No. 4.1e-10;  
 Matches 212; Conservative 85; Mismatches 382; Indels 400; Gaps 54;

23 PRLRLLMAGTARQV-----OGTGPRLACKES-----EYHYETACDSTGSRW 68  
 130 PSSNLTLLHAKADITYVRLKFTSRPESFAIKRREDGPMIPYIYSGSCENTYSKA 109  
 69 RVAVPHPPG-----LCTSLPDPVKTGBCSPSCNAGEFLDMKDKSCPCACGRRVSLGTGIR 123  
 190 NRGFIRGDEQOALCTD-----EESDF-----SPLTGVNAFSTP--- 224  
 124 FDEWDELPHGPAASLANNELDDSAESTGCTSSKWPREDGYIAPNDECTATIMAVNL 163  
 225 -----LEGHPSAYNFDNSPVLDEWYATD-----IRVTL 253  
 184 KOSGTVPNE-----YYPDSITFEFVONDOCOPNADSRMMKTEKGMEFHSV 233  
 254 NRLNTGDEVENDKVLKSYTAISD-----FVNGCKCKNGHASECKMKN-----EFDKL 303  
 234 ELNRGNVNL-----YWRTPAFSVMTKVPKPYLVNIAITGVAYTSECPG 278  
 304 VCNCKHNTYGVDCCKLPFENDRPMRATAES-----ASECLPC 342  
 279 -----KPGTY-----ADKSSPCKLCPCANSY-----SNKGETSCHQD 312  
 343 DGNRSGCYFDPPELYRSTGHGCTNCOQMDTGAHCERKXENFRLGNACSSCH-CS 401  
 313 P-----DKYSEKSSSCNWRPA-----CTDKDYFTYHT-----ACDANGETQLM 351  
 402 PVGSLSTQCDYGR-----CSCKPGVMGDKDRQOPGHSHTLEAGRCPCSDPGS1DEC 456  
 352 YKMAKPKICSEDLBGAVALPASGVKTHCPNPGCFEFTNNTST---COPCPGYSTSGNSDC 408  
 457 NVETGRVCCKNVEG-----FNCERCKPGFFLESSNPRGCTPC--FCGHSYSVC 504  
 409 TRCPAGTPAYGEEKKMMNLTPLTMMETTVLSGINFEKGMGWAVAGHITAGASDND 468  
 505 TN-----AVGYSV-----VISTSTFOIDEDGMRAEDRDSSEASLEMSEROD 546  
 469 FMILT-----LVYGFRRPOSVADTENEVARITFEVET---LGSVNCLEYMNGVN 518  
 547 IAVISDSFPRFRLAPAKFLKQVLYSTQN-----LSSFFVDRDRRLSMDLVLSGAG 601  
 519 SRTNTPVETWKG-S-KGOSYIIIEENTTSFTW-----AFQ-----RTFEH 559  
 602 LRVSVPILAGNSYSEFTTKYVFLHEATDYPMRPAITPPEFOKLNLNLSIKIRGYS 661  
 560 EASRKYTDVAKI-----YKIN-----YSIN 576  
 662 EESAGYLDVYTLASARPGVPATWVESCTCPVGGOFCCEMCLSGYRRETPNLAGPSPC 721  
 577 VTNVANGVASYCRP-----CALEASDVGSSTSCAPAGYIDRDSGT---CHSCP--PNTILK 628  
 722 VLCAENGHSFCDPEFTGVCNCRDNTAGPHCKGCSGXYIGDSTAGTSSQCGPCPGSSC 781  
 639 AHQPYGOV-ACVPCGPGTKNKKHSLC---YNDCTESNTPTR-----TFNTNFSALA 677  
 782 AVVPKTEKVCNCPGTGTGKRC-ELCDDGYFGDPLGRNGHVRRLCLOCSDNIDPNVAVG 840  
 678 NNTVLAGSPSTSGKLKTFHH-----FTLSLGNQGRKMSVCDNTVTLRIP 724  
 841 NENRLTG-----ECLKCTIYNAGYCDRCRKGDFGFLNPLADNADKACACN-----P 888  
 725 BEESGFSKITAYVCAVYIPEVTVYKAGV-----SSOPYSLAD-RLIGVYTD- 773  
 889 YETMGOSSCNVPTGCGECL-PIHYTGOCGACDPGFYMLGSGGGERCDHALGSTNGC 947  
 774 -----TLGDTSPALFHLIESIGIDVYIFFRSNDVYOSGSSGRTTIRVCSPKT 825  
 948 DIRTGOCCECPGITGO-----HCERC-----EVNHFGFGPGRCRCD-----CHPR- 988

QY 826 VPSLL-----DPTCS-----DGTDCGCFHFL-----WESAACPLC-----SVADY 864  
 DB 989 --GSLSLQCKDRCRCRGRGFGVGNRCDQCEHNFYRSMPPGQECPCRYLKVADH 1045

## RESULT 2

LMG1 MOUSE STANDARD; PRT: 1607 AA.  
 AC P02468;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Laminin gamma-1 chain precursor (Laminin B2 chain).  
 GN LAMC1 OR LAMC-1 OR LAMB-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88059118; PubMed=3680290;  
 RA Sasaki M., Yamada Y.;  
 RT "The laminin B2 chain has a multidomain structure homologous to the B1 chain."  
 RL J. Biol. Chem. 262:17111-17117(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89000737; PubMed=3167041;  
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;  
 RT "Primary structure of the mouse laminin B2 chain and comparison with laminin B1."  
 RL Biochemistry 27:5198-5204(1988).  
 RN [3]  
 RP SEQUENCE OF 1-239 FROM N.A.  
 RX MEDLINE=88228071; PubMed=2836421;  
 RA Ogawa K., Burdello P.D., Sasaki M., Yamada Y.;  
 RT "The laminin B2 chain promoter contains unique repeat sequences and is active in transient transfection."  
 RL J. Biol. Chem. 263:8384-8389(1988).  
 RN [4]  
 RP SEQUENCE OF 1391-1607 FROM N.A.  
 RX MEDLINE=85051302; PubMed=6209134;  
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil alpha-helix."  
 RL EMBO J. 3:2355-2362(1984).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.  
 RX MEDLINE=96196435; PubMed=8648630;  
 RA Steinfeld J., Mayer U., Timpl R., Huber R.;  
 RT "Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin gamma1 chain harboring the nidogen binding site."  
 RL J. Mol. Biol. 257:644-657(1996).  
 RN [6]  
 RP STRUCTURE BY NMR OF 824-881.  
 RX MEDLINE=96196435; PubMed=8648631;  
 RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R., Timpl R., Holak T.A.;  
 RT "Structure of the nidogen binding LE module of the laminin gamma1 chain in solution."  
 RL J. Mol. Biol. 257:658-668(1996).  
 CC -I- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.  
 CC -I- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.  
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),

CC	LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
CC	-I- SUBCELLULAR LOCATION: Extracellular.
CC	-I- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).
CC	-I- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC	-I- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC	-I- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC	-I- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
CC	-I- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
DR	EMBL; X05211; CAA28838.1; .
DR	EMBL; J03484; AAA39405.1; .
DR	EMBL; J02930; AAA39408.1; .
DR	EMBL; J03749; AAA39409.1; .
DR	PIR; A28469; MMSB2.
DR	PDB; IKLO; 20-AUG-97.
DR	PDB; ITLE; 12-FEB-97.
DR	MGI; MG1:99914; LamC1.
DR	InterPro: IPR004089; Chmtaxis_transd.
DR	InterPro: IPR000561; EGF-like.
DR	InterPro: IPR001886; LamNT.
DR	InterPro: IPR000034; Laminin_B.
DR	InterPro: IPR002049; Laminin_EGF.
DR	Pfam; PF00052; laminin_B; 1.
DR	Pfam; PF00053; laminin_EGF; 10.
DR	Pfam; PF00055; laminin_Nterm; 1.
DR	PRINTS; PR00011; EGFLAMININ.
DR	ProDom; PD002082; LamNT; 1.
DR	ProDom; PD003031; Laminin_B; 1.
DR	SMART; SM00180; EGF_Lam; 9.
DR	SMART; SM00001; EGF_like; 1.
DR	SMART; SM00281; Lamb; 1.
DR	SMART; SM00136; LamNT; 1.
DR	PROSITE; PS00022; EGF_1; 8.
DR	PROSITE; PS01186; EGF_2; 2.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 10.
KW	Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
KV	SIGNAL
FT	1 33
FT	CHAIN
FT	34 1607 LAMININ GAMMA-1 CHAIN.
FT	34 283 LAMININ N-TERMINAL (DOMAIN VI).
FT	284 339 LAMININ EGF-LIKE 1.
FT	340 395 LAMININ EGF-LIKE 2.
FT	396 442 LAMININ EGF-LIKE 3.
FT	443 492 LAMININ EGF-LIKE 4.
FT	493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	503 687 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	688 721 LAMININ DOMAIN IV.
FT	722 770 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT	771 825 LAMININ EGF-LIKE 6.
FT	826 881 LAMININ EGF-LIKE 7.
FT	882 932 LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).
FT	933 980 LAMININ EGF-LIKE 9.
FT	981 1028 LAMININ EGF-LIKE 10.
FT	1029 1607 LAMININ EGF-LIKE 11.
FT	DOMAIN
FT	1034 1594 DOMAIN II AND I.
FT	COILED COIL (POTENTIAL).
FT	DISULFID
FT	340 349 BY SIMILARITY.
FT	342 367 BY SIMILARITY.
FT	368 377 BY SIMILARITY.
FT	DISULFID
FT	380 393 BY SIMILARITY.
FT	DISULFID
FT	396 408 BY SIMILARITY.
FT	DISULFID
FT	398 414 BY SIMILARITY.
FT	DISULFID
FT	416 425 BY SIMILARITY.
FT	DISULFID
FT	428 440 BY SIMILARITY.

FT	DISULFID	443	454	BY SIMILARITY.
FT	DISULFID	443	461	BY SIMILARITY.
FT	DISULFID	445	472	BY SIMILARITY.
FT	DISULFID	475	490	BY SIMILARITY.
FT	DISULFID	722	731	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	752	768	BY SIMILARITY.
FT	DISULFID	771	779	
FT	DISULFID	773	790	
FT	DISULFID	793	802	
FT	DISULFID	805	823	
FT	DISULFID	826	840	
FT	DISULFID	828	847	
FT	DISULFID	850	859	
FT	DISULFID	862	879	
FT	DISULFID	882	896	
FT	DISULFID	884	903	
FT	DISULFID	905	914	
FT	DISULFID	917	930	
FT	DISULFID	933	945	BY SIMILARITY.
FT	DISULFID	953	952	BY SIMILARITY.
FT	DISULFID	954	963	BY SIMILARITY.
FT	DISULFID	966	978	BY SIMILARITY.
FT	DISULFID	981	993	BY SIMILARITY.
FT	DISULFID	983	999	BY SIMILARITY.
FT	DISULFID	1001	1010	BY SIMILARITY.
FT	DISULFID	1013	1026	BY SIMILARITY.
FT	DISULFID	1029	1029	BY SIMILARITY.
FT	DISULFID	1032	1032	INTERCHAIN (PROBABLE).
FT	DISULFID	1598	1598	INTERCHAIN (WITH CHAIN BETA-1).
FT	CARBOHYD	58	58	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	132	132	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	574	574	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	648	648	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1020	1020	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1105	1105	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1159	1159	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1173	1173	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1203	1203	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1221	1221	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1393	1393	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD	1437	1437	N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CONFLICT	216	216	G -> A (IN REF. 3).
FT	CONFLICT	260	260	E -> D (IN REF. 2).
FT	CONFLICT	337	337	S -> C (IN REF. 2).
FT	CONFLICT	447	448	LR -> PS (IN REF. 2).
FT	CONFLICT	544	544	D -> Y (IN REF. 2).
FT	CONFLICT	662	662	T -> S (IN REF. 2).
FT	CONFLICT	886	886	MISSING (IN REF. 2).
FT	CONFLICT	1158	1158	MISSING (IN REF. 2).
FT	CONFLICT	1434	1434	V -> A (IN REF. 2).
FT	CONFLICT	1475	1475	R -> K (IN REF. 4).
FT	CONFLICT	1576	1576	D -> N (IN REF. 4).
SQ	SEQUENCE	1607 AA;	177297 MW;	8187B08E4869EF242 CRC64;
Query Match                    4.1%;     Score 225;   DB 1;   Length 1607;				
Best Local Similarity        19.6%;   Pred. No. 3.le-08;				
Matches 212; Conservative    83; Mismatches 368; Indels 416; Gaps      53;				
OY	23 PLYMLLMWAGTAFVYT-----OQTGPELHMKCKS-----EYHVEYTCADSTGSRW	68		
DB	128 FNSIMLTLLKKAFFITTYRLKFHTSRPSFAIYKRTREDGWIPYOYSSGCEWTYSKA	187		
OY	69 RVAVHPBG-----LCTSLPPDFVKTECSFCSCNAGEFLDMKDQSKPCAEGRYSLGTGR	123		
DB	188 NRGFIRTSGDEQQALCTD-----EFSOI-----SPLTGGNVAVST---	222		
OY	124 FDEWDELPHGFASLSANMELDDSAESTGNCITSSKWVPREGDIANTBDCAATLMTAYANVL	183		
DB	223 -----LEGRPSAANFNPNSSPYLOEWVATD-----IRVTL	251		

OY 184 KOSGVNFE-----YYPPDSIIFFEFYVNDCCQPNADSKRMKTEKMEFHSV 233  
 DB 252 NRIINTEGDEVEPEPKVLSYIAISD-----FAVGRCCKGNHASECVKN-----EFDL 301  
 OY 234 ELNRGNVL-----YWRRTAFSVWTKVP-----KPVLENIAI 266  
 DB 302 MCCKINITYGVDCCKLPFENDRPWRRAAASASESLPCDCKNGSQRCYFDPPELYRS--- 358  
 OY 267 TGVATYSECPCKPRTYADKQSSFCCLCPANSYSNKGESCHOCDDPKSEKSSS--- 323  
 DB 359 TG--HGHCCTNCNDNTDGA-----CERCREN--FFRLGNTA--CSPCHCSPPVSLSTOC 408  
 OY 324 -----CNVRA-----CTDKDYFYTH-----ACDANGETOIMYMAKPKICSEDL 365  
 DB 409 DSYGRCSCKPGVMGDKDCRCQPFHSLTAGCRPCSCDLRGSTDECNVETRCYCKKDNVE 468  
 OY 366 GAVKLPAAGVTKHCPNCPNGFFKTNST---CQPCPYGSYNSGSDCTCGAGTEPVGFE 422  
 DB 469 G-----FNCERCKPGFFNLSSNPKGCTPC--FCFGSSVCTN-----AVGYS 509  
 OY 423 -YKWNNTLPYMETTVLSGINFEKMTGMEVAGDHITAAASDNDFMILTLYVPGFRP 481  
 DB 510 VYDISTFOIDED-----GMRV-----EQRDGESEASLEMSSDQ 543  
 OY 482 POSVADT--ENKEVARITFEVETLCSVNCLEF-----MGVNSRTN 522  
 DB 544 DIAVISDSYFRPYFLAPVKFLGNOLVLSYGNILSFSFVNRDRRLSLADLVLEGAGLRVS 603  
 OY 523 TPVETWKS--KQKOSYTYIIEENTTSFTW-----AFQ-----RTTPEASR 563  
 DB 604 VPLIAGNSYPSSETVYKTLFLHEATDPMRPAISPFEOKLNLNLSIKIRGYSERPA 663  
 OY 564 KYNDVA-----KIYSINVTN 580  
 DB 664 GYLDVTVLOSARPGVPATWESCTCPYVGGOCEFTCLDGRREPSLGPYSPVLCIT 723  
 OY 581 MNGVASTCRP-----CALBASDVSSCTSCSPAGYIYDRSGT---CHSCP-----P 623  
 DB 724 CNGHSETCDPBTGYCDRCRNTAGPHCEKSDGYGDSLTIGTSSDQPCPGSSCAIVP 783  
 OY 624 NT--ILKAHOPYGV--QACVPCPGPKNNKINSLCNDCTFSRNPTR-----TFNYS 672  
 DB 784 KTKEVVCHCTGTAGRCCELCDGYFGDPLGS-----NCPVRLCRPCQCNMD 833  
 OY 673 PSALANTYTLAGSPSTSGKLYFNH-----FTYSLCNGCRKMSVCTDWT 719  
 DB 834 PNAVNCNRLTG-----ECLKCIYNAGFYCDRCKEGFNGNPLADNPAKCKACACN-- 885  
 OY 720 DLRIEG--ESGFSKITAYVCAVILPEPVGYKAGV-----SSOPYSLADRLI 767  
 DB 886 ----PYGVVOQSSCNPTTGO--CQCL--PWSGRDGTCDPCYVYNQSGQGERCDCHA 937  
 OY 768 GVTDTMTDGTSPALFHLIESLIGPIVIFYRSNDVYOSGSSGRITTVRC----- 820  
 DB 938 LGSTNOCODIRIGCE-----CQPTGQHCERKETHNFG 973  
 OY 821 SPQKTVP-----GSLLL-----PGTCS-----DGTCDGCFNHL-----MESAACPLC 859  
 DB 974 GRECKRCDCDHGSLSLCKDDGRCREGRGVNKRCCOCEENFYNNSPMGCECPAC 1032  
 RESULT 3  
 LMG1\_DROME STANDARD: PRT: 1639 AA.  
 AC P15215; 024373; 09VTT18;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Laminin gamma-1 chain precursor (Laminin B2 chain).  
 GN LAM2 OR LAMC1 OR LAMG1 OR CG3322.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Canton-S, and Oregon-R;  
 RX MEDLINE=9129101; PubMed=1840513;  
 RA Chi H.-C., Jumanaga D., Wang S.Y., Hui C.-F.;  
 RT "Structure of the Drosophila gene for the laminin B2 chain.";  
 RL DNA Cell Biol. 10:451-466(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Oregon-R;  
 RX MEDLINE=8910916; PubMed=2912972;  
 RA Chi H.-C., Hui C.-F.;  
 RT "Primary structure of the Drosophila laminin B2 chain and comparison  
 with human, mouse, and Drosophila laminin B1 and B2 chains.";  
 RL J. Biol. Chem. 264:1545-1550(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90037237; PubMed=2808533;  
 RA Montell D.J., Goodman C.S.;  
 RT "Drosophila laminin: sequence of B2 subunit and expression of all  
 three subunits during embryogenesis.";  
 RL J. Cell Biol. 109:2441-2453(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanaides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Branton R.C., Rogers Y.H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abtill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Beldarr D., Bolshakov S.,  
 RA Beeson K.Y., Benos P.V., Berman J., Brokstein P., Brotler P.,  
 RA Borokova D., Botchan M.R., Bouck J., Cadieu E., Center A., Chandra I.,  
 RA Burtils K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA de Pablos K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foutin K.J., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Guan P., Harris M.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Idegami C.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Kenison J.A., Ketchum K.A.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Lai Z.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp Y., Lin X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Modarres C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacלב J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington I., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock K.C., Wu D., Yang S., Yao Q.A.,  
 RA Williams S.M., Woodage T., Worley K.C., Xu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [5]  
 RP SEQUENCE OF 344-1639 FROM N.A.  
 RC STRAIN-Oregon-R;  
 RX MEDLINE=88303364; PubMed=3405777;  
 RA Chi H.-C., Hui C.-F.;  
 RT "cDNA and amino acid sequences of Drosophila laminin B2 chain.";

```
CC CC      us thought to mediate the attachment, migration, and organization
CC CC      of cells into tissues during embryonic development by interacting
CC CC      with other extracellular matrix components.
CC CC      -I SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC CC      different polypeptide chains (alpha, beta, gamma), which are bound
CC CC      to each other by disulfide bonds into a cross-shaped molecule
CC CC      comprising one long and three short arms with globules at each
CC CC      end.
CC CC      -I SUBCELLULAR LOCATION: BASEMENT MEMBRANES (MAJOR COMPONENT).
CC CC      -I DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC CC      WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC CC      -I DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC CC      -I SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC CC      -I SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
CC CC      -I SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC CC      -----
CC CC      THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION
CC CC      BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC CC      AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC CC      USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC CC      MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC CC      ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sdb.ch/announce/
CC CC      OR SEND AN EMAIL TO license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: M58417; AAA28665.1; -.
CC CC      DR EMBL: M25063; AAA28664.1; -.
CC CC      DR EMBL: AE003551; AAP50238.1; -.
CC CC      DR EMBL: X07806; CAA30665.1; -.
CC CC      DR PIR: A31483; MMPB2.
CC CC      DR HSSP: P02468; 1TFE.
CC CC      DR FLYBASE: FBgn0002528; Laub2.
CC CC      DR InterPro: IPR0005561; EGF_1like.
CC CC      DR InterPro: IPR001886; LamNT.
CC CC      DR InterPro: IPR000034; Laminin_B.
CC CC      DR InterPro: IPR002049; Laminin_EGF.
CC CC      DR Pfam: PF00052; Laminin_B_1.
CC CC      DR Pfam: PF00053; Laminin_EGF_10.
CC CC      DR Pfam: PF00055; Laminin_Nterm_1.
CC CC      DR ProDom: PD002082; LamNT_1.
CC CC      DR ProDom: PD003031; Laminin_B_1.
CC CC      DR SMART: SM00180; EGF_Lam; 10.
CC CC      DR SMART: SM00281; EGF_1like; 1.
CC CC      DR SMART: SM00362; LamNT_1.
CC CC      DR PROSITE: PS00022; EGF_1; 8.
CC CC      DR PROSITE: PS01186; EGF_2; 1.
CC CC      DR PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
CC CC      KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC CC      LM Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC CC      FT STGNLU 1 33
CC CC      FT CHAIN 34 1639
CC CC      FT DOMAIN 34 298 LAMININ GAMMA-1 CHAIN.
CC CC      FT DOMAIN 358 358 LAMININ N-TERMINAL (DOMAIN VI).
CC CC      FT DOMAIN 359 413 LAMININ EGF-LIKE 1.
CC CC      FT DOMAIN 414 460 LAMININ EGF-LIKE 2.
CC CC      FT DOMAIN 461 513 LAMININ EGF-LIKE 3.
CC CC      FT DOMAIN 514 523 LAMININ EGF-LIKE 4.
CC CC      FT DOMAIN 524 709 LAMININ EGF-LIKE 5 (N-TERMINAL).
CC CC      FT DOMAIN 710 743 LAMININ DOMAIN IV.
CC CC      FT DOMAIN 744 792 LAMININ EGF-LIKE 6.
CC CC      FT DOMAIN 793 846 LAMININ EGF-LIKE 7.
CC CC      FT DOMAIN 847 901 LAMININ EGF-LIKE 8.
CC CC      FT DOMAIN 902 955 LAMININ EGF-LIKE 9.
CC CC      FT DOMAIN 956 1003 LAMININ EGF-LIKE 10.
CC CC      FT DOMAIN 1004 1049 LAMININ EGF-LIKE 11.
CC CC      FT DOMAIN 1050 1109 LAMININ II AND I.
CC CC      FT DOMAIN 1087 1247 COILED COIL (POTENTIAL).
CC CC      FT DOMAIN 1144 1247 COILED COIL (POTENTIAL).
CC CC      FT DOMAIN 1306 1627 COILED COIL (POTENTIAL).
CC CC      FT DISULFID 299 308 BY SIMILARITY.
CC CC      FT DISULFID 301 322 BY SIMILARITY.
```

FT	DISULFID	324	333	BY SIMILARITY.
FT	DISULFID	336	356	BY SIMILARITY.
FT	DISULFID	359	368	BY SIMILARITY.
FT	DISULFID	361	384	BY SIMILARITY.
FT	DISULFID	387	396	BY SIMILARITY.
FT	DISULFID	399	411	BY SIMILARITY.
FT	DISULFID	414	426	BY SIMILARITY.
FT	DISULFID	416	432	BY SIMILARITY.
FT	DISULFID	434	443	BY SIMILARITY.
FT	DISULFID	446	458	BY SIMILARITY.
FT	DISULFID	461	475	BY SIMILARITY.
FT	DISULFID	463	482	BY SIMILARITY.
FT	DISULFID	484	492	BY SIMILARITY.
FT	DISULFID	496	511	BY SIMILARITY.
FT	DISULFID	744	753	BY SIMILARITY.
FT	DISULFID	746	760	BY SIMILARITY.
FT	DISULFID	762	771	BY SIMILARITY.
FT	DISULFID	774	790	BY SIMILARITY.
FT	DISULFID	793	801	BY SIMILARITY.
FT	DISULFID	795	811	BY SIMILARITY.
FT	DISULFID	814	823	BY SIMILARITY.
FT	DISULFID	826	844	BY SIMILARITY.
FT	DISULFID	847	861	BY SIMILARITY.
FT	DISULFID	849	868	BY SIMILARITY.
FT	DISULFID	871	880	BY SIMILARITY.
FT	DISULFID	883	899	BY SIMILARITY.
FT	DISULFID	902	919	BY SIMILARITY.
FT	DISULFID	924	926	BY SIMILARITY.
FT	DISULFID	928	937	BY SIMILARITY.
FT	DISULFID	940	953	BY SIMILARITY.
FT	DISULFID	956	968	BY SIMILARITY.
FT	DISULFID	978	975	BY SIMILARITY.
FT	DISULFID	977	986	BY SIMILARITY.
FT	DISULFID	989	1001	BY SIMILARITY.
FT	DISULFID	1004	1016	BY SIMILARITY.
FT	DISULFID	1006	1022	BY SIMILARITY.
FT	DISULFID	1024	1033	BY SIMILARITY.
FT	DISULFID	1036	1047	BY SIMILARITY.
FT	DISULFID	1050	1050	BY SIMILARITY.
FT	DISULFID	1053	1053	INTERCHAIN (PROBABLE).
FT	DISULFID	1631	1631	INTERCHAIN (PROBABLE).
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	376	376	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	669	669	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	862	862	N-LINKED (GLCNAC. . .)

Query Match	4.18;	Score 224.5;	DB 1;	Length 1639;
Best Local Similarity	18.68;	Pident. No. 3.5e-08;		
Matches 191;	Conservative 98;	Mismatched 342;	Indels 395;	Gaps 52;
QY	34	TATVOTGTGTDEL-----	HACKESEHYEYTCDDSGSRWRRAVPHHTGLCTSL--	-PD 84
Db	174	TIYKRIKSESGPWPJYOFYSATCRDT-----	YSLPDBRAIRKKEGEAH--	ALCTSEYSDIS 226
QY	85	PVYGTSECFSCNAGEFLDMKDQCKRCACAGRRSLGTGIRFDEMDLPHGFASLSANMELD		
Db	227	PLNDGELARS-----	TEIGRPS--	GINFERSGELOEWATATDRIITLD 267
QY	145	-----	DSAAESTGNCT--SSKWP--	RGDIYAENPTDECT 174
Db	268	RLNTEGDELEFGDSOVLKSYFAISDIANGARCKGMHAKCPSTGMGE-----		
QY	175	ATLWAAVNLKQSGTANFEIYYPDSSITIEFFVYVNDQCPNADDSRMKMTTFKCGWEPHSVE		
Db	318	RLTVGECRHNDG-----	PDG--	DRCLPLVNDLKMKRST--STE 352
QY	235	LN-----	RGNNVL-----	YMTTAFSV 251
Db	353	VNEKACACNGLADKCFPDANLFNRHTGGHGLCDCKRNNRDGPNCCERCKENFYMRDGYCV		
QY	252	WTKYPRPLVAVNIIITGVATYS-EC-----	FPKPGTYADQKSSFCCKLCPANYSNKG	305
Db	413	-----	NCACDPVGSRSLOQCSHGKCKCPGVGTGDK-----	CDKCD--NNYIQEPR 455



306 TSCCHCDPDKYSEKSSSCVPRACDIDKDYFTHHTACDNDQMLMYKAKKICSEDL 365  
 456 HGCQCCGDC---SGSHQWT-PAC-----DTEGLICF-----CKENYE 489  
 366 GAVKLPASGVKTHCPNCPGPF---KTNSTCCPCPYGSGNSGSDCTRCPCAGTEPAV--- 419  
 490 G-----RRCNECKPGFENLDKNNRRCPCPC--FCIGHTSECMTAPGYSIVSVTSN 537  
 420 --GFEYKMMNTLPNTMETTVLSINFEYKGMTGWEVAG--DHITYAAGASDNDMILLTV 476  
 538 FNKEFERMTAADLNDREVDIKYN---QXSISIGTTAGNENHY----- 577  
 477 PGFRPPOSVMAD---TENKEVARITFEYFELICSNCLCYMVCNSTRNTPVETWKS-- 531  
 578 --FQAPDRFLGDORASYNRDKFKIOLVGOVANGVSDVILEGAGRSISLPIRA-QGNGI 634  
 532 --KGASTYITIEENTTSFTWAFORTTEHESRRKYNDVAKI-----YSINVTVM 581  
 635 PDGKVEYTERLHEH--HDYOMPS-----QSARGFLISLNLTAIKIRATYSVOGEAIL 687  
 582 NGV-----ASYCRPCALASDVSSCTSCPCAGYIDRDSG-----TCH 619  
 688 DVEIQTARHGAAGHPATWIEQCTCEGYLQFCESCAPGRHSRPARCGPMPICPDCH 747  
 620 S-----CPPT--ILKAHQPYQACVPCGPGTKNNKIHSLCYNDCTSRNTPRTFVNF 673  
 748 GHADIDSETGRICQHNTHG--DNCDQCAKGFYGNALGG--TPNDC--KRCP----- 794  
 674 SALANTVTLAGPSPFYSKGLKYFHHTLSLGNORRMSVCTDNTVTLRIEBS----- 726  
 795 -----CPNDGACCIQINDPIYICRECPGFHSGRCE 824  
 727 --ESGESKSTTAYV-----CQAV-----IIPREYGVKAGVSOPVSLADRLIGVTTMTL 775  
 825 QCSDFEFGDPTGLIGEVQTCSCDCNGVNDPNAVNCNNTTGECLKIHNTAGEHCDQL 884  
 776 DG-ITSPAEFLHESLGIPIVIFFYRSNDVYOSCSSGSRSTTRVRCSPQKTPYSGLLPG 834  
 885 SGHGGDPLALPH---GRDRCSCYENGTEPDQSIITRCDQVYGCQCPNVIIGBDC--G 938  
 835 TCSDG-----TCDCGFHFLMESA 853  
 939 ECGGYFNIRISGNCENCILCDPVGSYNSTCDRYSGOCHCRPGVMGRCQCCENYFGEFS 998  
 854 AACPLC 859  
 999 EGCKPC 1004  
 RESULT 4  
 TSA4\_GIALA STANDARD; PRT; 713 AA.  
 ID TSA4\_GIALA  
 AC P21849.1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major surface-labeled trophozoite antigen 417 precursor.  
 GN TSA 417.  
 OS Giardia lamblia (Giardia intestinalis).  
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.  
 OX NCBI\_TaxID=5741;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 30957 / WB;  
 RX MEDLINE=90280395; PubMed=2352929;  
 RA Gillin F.D., Hagblom P., Harwood J., Alely S.B., Reiner D.S.,  
 RT McCaffery M., So M., Guinea D.G.;  
 RT "Isolation and expression of the gene for a major surface protein of  
 Giardia lamblia".  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).  
 RN [2]  
 RP SEQUENCE OF 480-620 FROM N.A.

RC STRAIN=AD-1;  
 RX MEDLINE=93314970; PubMed=8325510;  
 RA Ey P.L., Mayrhofer G.;  
 RT "Two genes encoding homologous 70-kDa surface proteins are present  
 within individual trophozoites of the binucleate protozoan parasite  
 Giardia intestinalis".  
 RL Gene 129:257-262(1993).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE  
 PLASMA MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.  
 CC -1- DOMAIN: CONTAINS 29 REPEATS OF THE CYXC MOTIF.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M3641; AAA02688.1; -;  
 CC EMBL: M97488; AAA02581.1; -;  
 CC PIR: A35502; A35502.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR005127; Giardia\_VSP.  
 CC Pfam: PF03302; VSP; 2.  
 CC SMART: SM00181; EGF; 1.  
 CC SMART: SM00261; FU; 3.  
 CC SMART: SM00261; FU; 3.  
 CC Signal: Antigen; Glycoprotein; Transmembrane; Repeat.  
 CC CHAIN 1 18 713  
 FT DOMAIN 18 679  
 FT TRANSMEM 680 708  
 FT DOMAIN 709 713  
 FT CARBOHYD 289 289  
 FT CARBOHYD 676 676  
 FT VARIANT 582 582  
 FT VARIANT 606 606  
 FT SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;  
 SO  
 Query Match 3.8%; Score 211.5; DB 1; Length 713;  
 Best Local Similarity 21.6%; Pred. No. 1; Le-07;  
 Matches 155; Conservative 72; Mismatches 278; Indels 211; Gaps 39;  
 4 PGHSHHLASRVGRTERIRIPRLMLMAGTARQVOTGTEPELHAKSESEYHETAC-D 62  
 99 PGHSLICLSSDDGDVCTEAPAGYFAPVGAANTQESV-----IACGD 138  
 63 STGSRWRAVPHPTGLC---TSLPDPVKTE-----CSFSCNAGEFLDMKDOCKPCA 112  
 139 TTGVITIAAGNTYKGIADCAECASAPDATAGAEKAVATCT-KGVSKYL-KONVCVKA 195  
 113 EGRVSLGIRPDEWDELPHGFASLSANMELIDSAESTGCTSSKVPKGVDIANTDTE 172  
 196 Q--CNSGSTNKFVANDSDENGNKVCSCDNLNGVA---NCDTCYSYDESKRTI---K 244  
 173 CT-ATLMAVAVNLKOSGYVNFYYPDSSIIIEFFVNDQCOCPN---ADDSRMKMTTEKGM 228  
 245 CTKCTDNNVLTCTTSEGT-----SCVADQCKDGEFFPDDSDS----- 279  
 229 EFHSELNKGNNVVLWRTTAESVMTKVPKPVLVNRIAL-----TGVAYTISECFPKPG 281  
 280 -----SAGNCKLPCNDSTDS-----IANCATCALVSGRSGAALVT-CSACTDG 321  
 282 --TYADR--QGSSFCCKLCPANSYNNKGETS--CHQCPDXYSEKSSSCVPRACDIDK 334  
 322 YKPSADKTCCEAVSNCKTGTGCKACSNSEKENVCTDCD---GSYLLTPTSGCID-- 372  
 335 YFTTHACANGPTQMLMYMAKPKICSEDLGCAVVLPAASGVKT-----HCPPCNPFPFKT 389

Db 373 -----SCAKIGNYGGATEGAK-KLCKE-----CTANCKTDDGQCCOACNDGFRK- 417  
 Qy 390 NNSTCOPCYG-----SYNSGSDCTRCPCPTGPAVGFEXKMMNLTPTMETTVLSGINFEY 445  
 Db 418 NDDACSPCHESKCTCSAGTASDCTCECPG-----KALRYGD 453  
 Qy 446 KGMTEWAGDHITTAAGASNDPMILTVPGF-----RPSVYADENK--EV 494  
 Db 454 DGTG--TCGEGCTTGACAGCTGGLIIGASVCSCTATTEYTPONGVAPARATPT 511  
 Qy 495 ARITVEFTLCSVNCLELYFW--GVNSRTNPEWT-----KSGKOSYTYIIEENTT 546  
 Db 512 CNDSPIONGCGTADNPFKMGCCYETVYPCKTVICISAPNGCTGCKADYKLDGSL 571  
 Qy 547 TSTWAFORTTFHESKRKYNDVAKIYSINVTVMNGV--ASYCPALF--ASDVS 600  
 Db 572 -----FVSCGCEKAS-----STCTCTCDEGYKASACACRKCASCETGNGAT 616  
 Qy 601 SCTSCPAGY-IBRDSCTCHSCPNTILKAHPYGVQACVPGSPGPKNNKHSICLY 655  
 Db 617 TCKACATGYKTAGSGEGACTSCSD-----SNGVTGKGLNCAP-PPNNKGSVLCY 667

RESULT 5  
 LM33\_MOUSE STANDARD; PRT: 1581 AA.  
 AC 09R0B6; 09RTW6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Laminin gamma-3 chain precursor (Laminin 12 gamma 3).  
 GN LM33.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBI\_TaxId=10090;  
 RP SEQUENCE FROM N.A.  
 RA Albus A.M., Burgeson B., Champliand M.-F., Koch M., Olson P.;  
 RT "Mouse laminin 12 gamma 3 chain."  
 RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-1526 FROM N.A.  
 RA MEDLINE-99253969; PubMed-10318827;  
 RT Iivanainen A., Morita T., Tryggvason K.;  
 RT "Molecular cloning and tissue-specific expression of a novel murine laminin gamma3 chain".  
 RL J. Biol. Chem. 274:14107-14111(1999).  
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.  
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.  
 CC THE GAMMA-3 CHAIN IS A SUBUNIT OF LAMININ-12.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: Strongly expressed in capillaries and arterioles of kidney as well as in interstitial Leydig cells of testis.  
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF083372; AAF08983.1; -  
 CC DR EMBL; AF079520; AAD29851.1; -  
 CC DR HSSP; P02468; IKLO.  
 CC DR MGD; MGI:1344394; lamc3.  
 CC DR InterPro; IPR000561; EGF-like.  
 CC DR InterPro; IPR001866; LAMNT.  
 CC DR InterPro; IPR000034; Laminin\_B.  
 CC DR InterPro; IPR002049; Laminin\_EGF.  
 CC DR Pfam; PF00052; Laminin\_B\_1.  
 CC DR Pfam; PF00053; Laminin\_EGF\_10.  
 CC DR Pfam; PF00055; Laminin\_Nterm\_1.  
 CC DR PRINTS; PR00011; EGF\_LAMININ.  
 CC DR PRODOM; PD002082; LAMNT\_1.  
 CC DR SMART; SM00180; EGF\_Lam; 10.  
 CC DR SMART; SM00281; Lamb; 1.  
 CC DR SMART; SM00136; LAMNT\_1.  
 CC DR PROSITE; PS00022; EGF\_1; UNKNOWN\_8.  
 CC DR PROSITE; PS01186; EGF\_2; 2.  
 CC DR PROSITE; PS01248; LAMININ\_TYPE\_EGF\_11.  
 CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 CC Laminin EGF-like domain; Cell adhesion; Repeat; signal.  
 CC FT SIGNAL 1 28  
 CC FT CHAIN 1 28  
 CC FT DOMAIN 29 1581 LAMININ GAMMA-3 CHAIN.  
 CC FT DOMAIN 29 279 LAMININ N-TERMINAL (DOMAIN VI).  
 CC FT DOMAIN 280 335 LAMININ EGF-LIKE 1.  
 CC FT DOMAIN 336 391 LAMININ EGF-LIKE 2.  
 CC FT DOMAIN 392 438 LAMININ EGF-LIKE 3.  
 CC FT DOMAIN 439 488 LAMININ EGF-LIKE 4.  
 CC FT DOMAIN 489 684 LAMININ EGF-LIKE 5 (N-TERMINAL).  
 CC FT DOMAIN 685 718 LAMININ DOMAIN IV.  
 CC FT DOMAIN 719 766 LAMININ EGF-LIKE 6.  
 CC FT DOMAIN 767 821 LAMININ EGF-LIKE 7.  
 CC FT DOMAIN 822 877 LAMININ EGF-LIKE 8.  
 CC FT DOMAIN 878 927 LAMININ EGF-LIKE 9.  
 CC FT DOMAIN 928 975 LAMININ EGF-LIKE 10.  
 CC FT DOMAIN 976 1024 LAMININ EGF-LIKE 11.  
 CC FT DOMAIN 1025 1581 LAMININ II AND I.  
 CC FT DOMAIN 1029 1046 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1112 1153 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1208 1231 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1438 1468 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1510 1575 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 128 128 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 304 304 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 337 337 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 640 640 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 849 849 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 991 991 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 1162 1162 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 1196 1196 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 1320 1320 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CARBOHYD 1514 1514 N-LINKED (GLCNAc... ) (POTENTIAL).  
 CC FT CONFLICT 9 9 L-> F (IN REF. 2).  
 CC FT CONFLICT 190 190 F-> T (IN REF. 2).  
 CC FT CONFLICT 195 195 R-> K (IN REF. 2).  
 CC FT CONFLICT 221 221 G-> S (IN REF. 2).  
 CC FT CONFLICT 394 394 C-> R (IN REF. 2).  
 CC FT CONFLICT 471 471 C-> Y (IN REF. 2).  
 CC FT CONFLICT 1150 1150 L-> Q (IN REF. 2).  
 CC FT CONFLICT 1387 1387 H-> Q (IN REF. 2).  
 CC FT CONFLICT 1438 1439 AS-> TI (IN REF. 2).  
 CC FT CONFLICT 1479 1479 V-> I (IN REF. 2).  
 CC SO SEQUENCE 1581 AA; 172316 MW; 51DFAD1F956AE81 CRC64;

Query Match 3.88; Score 209.5; DB 1; Length 1581;  
 Best Local Similarity 18.64; Pred. No. 4.1e-07;  
 Matches 218; Conservative 107; Mismatches 394; Indels 451; Gaps 60;



23 PRLRLIMAGTAFOYT-----OQTGPRHACK-----SEHYEYACDST----- 64  
 124 PTNVNLLSLGKAYELTYLAKFHTSRPESFATYKRYTASGHWMEYXYYSCKQYGRP 183  
 65 -GSMWRAVHTTGLCTS-LPD--PVKGTCEGSPCNAGEFLDMKDCKPCACREHYSLGT 120  
 184 EGHVLRGEEDERVAFTSESDISPLNGVAFS-----TLEGKPS-- 224  
 121 GIREDEMDELPHGPAASLANMELDSDAESTGCTSSKVVPRGDIYAFNTDECTATLMA 180  
 225 AVNFESVYL-----QENVTSID-LILSLDLNLT---FG 254  
 181 VNLKSGSYNVEYYPPDSSTIFFEYV-----ONDCCPNA----- 215  
 255 DDFKDPVLOSITY-----AVSDPSVGRCCKNCHHASECPNAGOLACRQHNTTGVDC 310  
 216 -----DSDRMKTKTER-----GMEFISVELANGNNVLYWRTTAFSVMTKVPVYL 260  
 311 ERLPLFPQDRPWARGTAEDANECLPCNCSHSECTDREL--YRST----- 355  
 261 VRNIAITGYAITSCEPCKPGTYADKQSSFCILPANSSYNNKETSCHOCDDPKYSENG 320  
 356 -----GHGHCORCR-----DHITGPHERCEKRYWMSPTKCPQPCD--CHPAG 398  
 321 SSSCNVPACTDKDYFYHTTACDANGETOAMKAKPKICESEDLGAVKLPAASGVKTHCP 380  
 399 SLSTIQ-----CDNSG-----VC-----PCKPVTGKMC--CD 422  
 381 PCNPFKTNNSCOP-----CPYGSYNSGDCRTPAGT----- 415  
 423 RCLGFFHLSBGRPCACNAGVSLGTCDBRSCNCKENYVGSJLDCRCRGTENLOPHN 482  
 416 -----EPAVGEYKMMNLPMTMETTVSGINFEKMGKTGEVAGDHI- 458  
 483 PVGSSCCCYGHSKYCSPAAGFO-----EHHITS--DFRH-GAGGQJIRSMGV 528  
 459 -----YTAG-----ASDNDFM-----LTLVVP--GFRPP----- 482  
 529 KRPIMOSGGLLIRGGEELSAKKFLGRLGRLSYGCPVILTLQVPPGSPPIQLRLG 588  
 483 -----OSVADT--ENEVARITEVEETLCSVNCLEFNV 515  
 589 AGIALSLRSSLPSPODTRORARRVOLLOETSESESPRPFHGRILLS----- 639  
 516 GVNSTRNTPEVEMKSGKOSYTYIIEENTTSFTMAFORTTHEAS-----BKYN 567  
 640 -----NLTAISTVSGGCHGGLVLCVOLTSAWP--QRELABPASWETCLCPQCTG 693  
 568 DVAKIYSINTVNM--NGVASYCRCALASDV-----GSSCTSCPAGY 610  
 694 QCFECALGKREIPHGVPANCTPCTCNOHGTCDPNTGILCGHHTGSPCERCMGIFY 753  
 611 IDRBSG--TCHSCP--PNTILKAHP--YGVQACVPCPGRETKNKHISLCYADCFESNTP 665  
 754 GNAPSRAADCOPPCPGGACATIPESGDVCTHCPQORRCEES--CED- 803  
 666 TRTFNYSALANTVTLAGSPFTSGIKYFHHFTLSLGN-----QGRMSVCTDNVTL 721  
 804 -----GFTDPLGLSQAQ--PCRCQCSGNVDLANAGNCDPHSGHCLR--CLYNTGA 853  
 722 RIBEGESFESITAYVCOAVIIPPEVTGYKAVSSOPS-----LADRLIGTDTMTLDG 777  
 854 HCHREGF-----YGSANATRPVCKAPCSGDLRGVSEKTCMP 893  
 778 ITSPAELHESLGPVY-----IFEYRSNDVYTOSSGSGSTTIRVCSQKTV 826  
 894 VTS-----QCVLPYVSGRDCSCSGFYDLDSGGGSCCHPLGSLJENKCHPK--- 943  
 827 PGLLLPGTC-----SDTDCGCFHFLMSAACPCLCSADVADNHAIVSCVAGIKTV 880  
 944 -----TGCPCRPVGTGQACDRCOLGFFGFSIKGCRRCRCSPLGAASSOC---HENSTC 994  
 881 VWRPILCSGGISLPEORVYICKTIDFWLKVQ--ISAGTCTAILTLVLCYFWKKNOKLEY 939

Db 995 VCR-----PGFVGYKDR-----COD-NFPLADBDTGCEQECP-----TCYALVKEBAKL 1038  
 QY 940 KYSKLVNATLK--DCDLPAADSCALMEGE 967  
 Db 1039 KARLMMEGWLQBRSDCGSPWG-PLDILQGE 1067  
 RESULT 6  
 PKC5 MOUSE STANDARD; PRT; 1877 AA.  
 ID AC 004592; 062040; (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 41, Last annotation update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
 DE (Proteinase convertase PC5) (Subtilisin/kexin-like protease PC5)  
 DE (convertase PC5) (PC5) (Subtilisin-like proprotein convertase 6)  
 DE (SPC6).  
 GN PCSK5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1] SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).  
 RP STRAIN=ICR; TISSUE=Intestine;  
 RC MEDLINE=93327934; PubMed=8335106;  
 RA Nakagawa T., Murakami K., Nakayama K.;  
 RA "Identification of an isoform with an extremely large Cys-rich region  
 of PC5, a kex2-like processing endoprotease.";  
 RL FEBS Lett. 327:165-171(1993).  
 RN [2] SEQUENCE FROM N.A. (ISOFORM PC5A).  
 RP TISSUE=Brain, and Intestine;  
 RC MEDLINE=93224489; PubMed=8468318;  
 RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,  
 RA Nakayama K.;  
 RA "Identification and functional expression of a new member of the  
 mammalian kex2-like processing endoprotease family: Its striking  
 structural similarity to PACED.";  
 RL J. Biochem. 113:132-135(1993).  
 RN [3] SEQUENCE FROM N.A. (ISOFORM PC5A).  
 RP TISSUE=Adrenal cortex;  
 RC MEDLINE=93342056; PubMed=8341687;  
 RA Lussan J., Vieu D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
 RA "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a  
 candidate proprotein convertase expressed in endocrine and  
 nonendocrine cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
 RN [4] PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
 RP MEDLINE=97103178; PubMed=8947550;  
 RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,  
 RA "The isoforms of proprotein convertase PC5 are sorted to different  
 subcellular compartments.";  
 RL J. Cell Biol. 135:1261-1275(1996).  
 RN [5] DEVELOPMENTAL EXPRESSION.  
 RP MEDLINE=96293359; PubMed=8698813;  
 RA Constam D.B., Calton M., Robertson E.J.;  
 RA "Spcc, Spcc, and the novel protease SPC7 are coexpressed with bone  
 morphogenetic proteins at distinct sites during embryogenesis.";  
 RL J. Cell Biol. 134:181-191(1996).  
 RN [6] DEVELOPMENTAL EXPRESSION.  
 RP MEDLINE=97436919; PubMed=9291583;  
 RA Rancourt S.L., Rancourt D.E.;  
 RA "Murine subtilisin-like proteinase SPC6 is expressed during embryonic  
 implantation, somitogenesis, and skeletal formation.";  
 RL Dev. Genet. 21:75-81(1997).

CC -1 FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY  
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE  
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE  
CC FOR THE MAURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED  
CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION  
CC OF GROWTH FACTORS.  
CC CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR  
CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-YAA BONDS, WHERE XAA  
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.  
CC SUBCELLULAR LOCATION: PCS5 IS SECRETED THROUGH THE REGULATED  
CC SECRETORY PATHWAY. PCS5 IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO  
CC A PARACELLULAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH  
CC EARLY ENDOSOMES.  
CC -1 ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS. PCS5/LONG (SHOWN HERE)  
CC AND PCS5/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1 TISSUE SPECIFICITY: PCS5 IS EXPRESSED IN MOST TISSUES BUT IS MOST  
CC ABUNDANT IN THE INTESTINE AND ADRENALS. PCS5 IS EXPRESSED IN THE  
CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.  
CC -1 DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,  
CC EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE KIDNEY AND THE LIVER,  
CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT  
CC E7.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT  
CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND  
CC NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK  
CC EG.5 AND E11.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK  
CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED  
CC TO THE CONDENSING MESENCHYMA SURROUNDING THE CARTILAGE. AT THIS  
CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL  
CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,  
CC ISOPORM A IS MOST ABUNDANT IN EPITHELIAL CELLS OF THE INTESTINAL VILUS.  
CC ISOPORM B OCCUR AT E12.5.  
CC -1 DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE  
CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC  
CC RETICULUM.  
CC -1 DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN  
CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS  
CC WITH THE TGN SORTING PROTEIN PACS-1.  
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
CC -1 SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as their content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
CC -----  
CC EMBL: D17583; BAA04507.1; -  
CC EMBL: D12619; BAA02143.1; -  
CC EMBL: L14932; AAA47636.1; -  
CC PIR: JX0248; JX0248.  
CC PIR: A48225; A48225.  
CC HSP: G99405; IMPP.  
CC MEROPS: S08.076; -  
CC MGD: MGI:97515; PCS5.  
CC InterPro: IPR000561; EGF-like.  
CC InterPro: IPR002174; Furin-like.  
CC InterPro: IPR002884; P\_domain.  
CC Pfam: PF00082; Peptidase\_S8.  
CC Pfam: PF01483; P\_PARTIAL.  
CC PRINTS: PR00723; SUBTILISIN.  
CC ProDom: PD000717; P\_domain; 1.  
CC SMART: SM00181; EGF\_3.  
CC SMART: SM00001; EGF-like; 2.  
CC SMART: SM00261; FU\_22.  
CC PROSITE: PS00136; SUBTILASE\_ASP; 1.  
CC PROSITE: PS00137; SUBTILASE\_HIS; 1.  
CC PROSITE: PS00138; SUBTILASE\_SER; 1.  
CC Hydrolase: Serine protease; Glycoprotein; Zymogen; Signal;

FW	KW	Match	Score	DB 1	Length	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB 400	DB 401	DB 402	DB 403	DB 404	DB 405	DB 406	DB 407	DB 408	DB 409	DB 410	DB 411	DB 412	DB 413	DB 414	DB 415	DB 416	
----	----	-------	-------	------	--------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--





```

RP SREUNCEFROM N.A.
RC STRAIN-Bristol N2.
RA Mux P.,
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U46673; AAC48152.1; -.
CC HSP: P02468; ILE.
DR WormPep: C54D1.5; CE06981.
DR InterPro: IPR001886; LAMNT.
DR InterPro: IPR000034; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00052; Laminin_B; 1.
DR Pfam: PF00053; Laminin_EGF; 10.
DR Pfam: PF00055; Laminin_Nterm; 1.
DR PRINTS: PR00011; EGFLAMININ.
DR ProDom: PD002082; LAMNT; 1.
DR ProDom: PD003031; Laminin_B; 1.
DR SMART: SM00180; EGF_Lam; 9.
DR SMART: SM00281; EGF_Like; 2.
DR SMART: SM00136; LAMNT; 1.
DR PROSITE: PS00022; EGF_1; 8.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
DR Hypothetical protein; Laminin_EGF-like domain; Signal; Repeat.
KW SIGNAL.
FT CHAIN ? 1 1557 LAMININ-LIKE PROTEIN C54D1.5.
FT DOMAIN ? 271 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 272 331 LAMININ EGF-LIKE 1.
FT DOMAIN 332 387 LAMININ EGF-LIKE 2.
FT DOMAIN 388 434 LAMININ EGF-LIKE 3.
FT DOMAIN 435 487 LAMININ EGF-LIKE 4.
FT DOMAIN 488 498 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 498 688 LAMININ EGF-LIKE 6 (C-TERMINAL).
FT DOMAIN 687 732 LAMININ EGF-LIKE 7 (INCOMPLETE).
FT DOMAIN 737 769 LAMININ EGF-LIKE 8.
FT DOMAIN 770 819 LAMININ EGF-LIKE 9.
FT DOMAIN 820 874 LAMININ EGF-LIKE 10.
FT DOMAIN 875 930 LAMININ EGF-LIKE 11.
FT DOMAIN 931 978 LAMININ EGF-LIKE 11.
FT DOMAIN 979 1025 LAMININ EGF-LIKE 11.
FT DISULFID 272 281 BY SIMILARITY.
FT DISULFID 274 295 BY SIMILARITY.
FT DISULFID 297 306 BY SIMILARITY.
FT DISULFID 309 329 BY SIMILARITY.
FT DISULFID 332 341 BY SIMILARITY.
FT DISULFID 334 357 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 372 385 BY SIMILARITY.
FT DISULFID 388 400 BY SIMILARITY.
FT DISULFID 390 406 BY SIMILARITY.
FT DISULFID 408 417 BY SIMILARITY.
FT DISULFID 420 432 BY SIMILARITY.
FT DISULFID 435 449 BY SIMILARITY.
FT DISULFID 437 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 470 485 BY SIMILARITY.
FT DISULFID 470 778 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 791 800 BY SIMILARITY.
FT DISULFID 803 817 BY SIMILARITY.

```

FT DISULFID 820 834 BY SIMILARITY.  
 FT DISULFID 822 841 BY SIMILARITY.  
 FT DISULFID 844 853 BY SIMILARITY.  
 FT DISULFID 856 872 BY SIMILARITY.  
 FT DISULFID 875 894 BY SIMILARITY.  
 FT DISULFID 902 901 BY SIMILARITY.  
 FT DISULFID 912 912 BY SIMILARITY.  
 FT DISULFID 915 928 BY SIMILARITY.  
 FT DISULFID 931 943 BY SIMILARITY.  
 FT DISULFID 933 950 BY SIMILARITY.  
 FT DISULFID 952 961 BY SIMILARITY.  
 FT DISULFID 964 976 BY SIMILARITY.  
 FT DISULFID 979 991 BY SIMILARITY.  
 FT DISULFID 981 998 BY SIMILARITY.  
 FT DISULFID 1000 1009 BY SIMILARITY.  
 FT DISULFID 1012 1023 BY SIMILARITY.  
 FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1168 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1211 1211 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1557 AA; 17723 MW; CAROBS1FBD5EBD2P CMC64;.

Query Match 3.58; Score 192; DB 1; Length 1557;  
 Best Local Similarity 19.38; Pred. No. 7.4e-06;  
 Matches 211; Conservative 99; Mismatches 390; Indels 394; Gaps 54;

20 RIPLMLLLMAGTAQVOTGTPELHACKESYHEXACOSTSKRRVAVPHTPGIC 79  
 114 QOYPTTNLTTLTKSFDITYRLKISPRESEFTYKTHHDSWEMEQF----- 164  
 80 TSLDPVKGTECSFSCNAGEFLDMKQSCPCAGRYSLGTGIRPDEMDLPHGFASLSA 139  
 165 -----YSGSCA-----TYGLS-----DRAPILPGNEALRQC 191  
 140 NMLDLSAESTGNCSTSSKWPGRDYIAFNTE-----CTATLMAVALKQSGVNEE- 192  
 192 TKERSDLSPTLGNIAFSTLEGRPSAHAESEVLOKWTASAI-RISLMMNTFGDEVE 250  
 193 -----YYPDSIIFFEFY-----ON----- 208  
 251 KDOVLRSYY-----ALSDFAVGRCKCNGHASECVSSSVDEGENRLVCRCEHNTGAGLC 306  
 209 DDCOPNADSRMKTTEKGEHVEYN-----RGNNVLYWRTPAFSVTKYKPYVL 260  
 307 NCELPFYNDRRRSGT-----SVEANECIACNCSQLSRKRCFFDOOLEE----- 350  
 261 VANITVAATSECFPCPKPTYADKQGSFCKLCRANSTSNKGETSCHQCDPKYSEKG 320  
 351 -----TG-HGHCIDCOGNT-----OG-VHCQCCIANMRREGENYCAVCS-----CNEIG 394  
 331 SSS-----CNVRPACTCK-----DYFY-----THACDANGEMQOLMKRAKAKT 359  
 395 SLSTQCNECKCKCKCKRGVGRFCDOCLDGFYSTNGCKNCCCEISGLN-----NDPRC 449  
 360 CSELEAVLPLASGVKTHCPNPGFF-----KTNNSTQPCPYGYSYSGSDCTRCR- 412  
 450 DSSSGSCSLANVEG-----ROCDKCKRGYFPLSTENOFCTPC-RCFGHSSICNTADGYFA 505  
 413 -----AGTEAVGEYKRWNTL-PTNMTETLSCINEFYGMGMVEYAGD 456  
 506 MANSVFFDDKKRWAG-QRRIGLQDOTQWALDKAVALSDTNSPYR-----VABE 555  
 457 HITTAGASDNDPMLTLTVPGFRPOSV-----MADTENKEVARI-----T 498

Db 556 QFLDQBSRYSNODLVFTLKAKHVTNODVKDIIIVGADQDELSTSTTAQGNPPTTEAQT 615  
 QY FVEETLSCVNCLEF-----AVGNSRNTVEVWKSSKQOSYTYIEENTTS 548  
 Db 616 YRFR-----VHADPYFGYPRINELDFIGLS-NTAIKIRCTYSYKDIGYLSNVNIGTA 669  
 QY 549 -----FTW-----AFORTTPEASRKYNTNVAIYISNV 577  
 Db 670 GVAPSANPKQATWIEHCECLPEFGVQFCESGSEFRRET-----KFGPFNHCICDC 723  
 QY 578 TNWANGVASCRCPCALASDVSSCTSCPAGYIIDROSGT-----CHSD-PN---TJLKAH 630  
 Db 724 HNSNSCEASGSCICEHNTAGDTCERCARGYBDALQAGTEEDQKCPCHNDPCILHAD 783  
 QY 631 -----QPYGQACVPCGCGTKNNKTHSLCYNDCTFSRNTPTRTFNYSALANTYT 681  
 Db 784 GDTCTECPNGYTGRCRCESDGYFGNPKDTECEVCACSGNTDP-----NSIGCDK 836  
 QY 682 LAG-----GSPFYSKGL-----KTFHFTSLCGN 706  
 Db 837 ITGECKKCIENTHGFNENCCKPGYWGDLIEPKNGOSCGCFAAGTRRRNNDYTLLECNO 896  
 QY 707 QGRKMSVCTDWTDLRIPEGESFSKSIYAVQAVIIP-----EV-TGY---KAG 754  
 Db 897 QDQGD-CLPNVIGIQDQCAHGFYNTISGLQGCQECNCDPLGSEGNCTCDVNTGQCCCKPG 955  
 QY 755 VSSQPSVLADRLIGVITDMLDITSPAEFLHESIGIPDVIFFRSNDYQSC--SSGR 812  
 Db 956 VTGQ-----KCDRCADYHFGFSANG-CQPCDEYIGS-----ENQOCDVNSG- 996  
 QY 813 STTIRVRCSEQKTPVPSLLPCTCSIDGTCDC--NFHLMESAACPLCSVADYHAIVSS 870  
 Db 997 -----QCLCKENVEGR-----KCDCAENRIGITQGLPC-----DD 1028  
 QY 871 CVAGIQKTYVWRE 884  
 Db 1029 CYTLIOSRVNVFRE 1042

RESULT 10  
 LM33\_HUMAN  
 ID LM33\_HUMAN STANDARD; PRT: 1587 AA.  
 AC Q9Y6N6;  
 DT 15-JUN-2002 (Ref. 41, Created)  
 DT 15-JUN-2002 (Ref. 41, Last sequence update)  
 DT 15-JUN-2002 (Ref. 41, Last annotation update)  
 DE Laminin gamma-3 chain precursor (Laminin 12 gamma 3).  
 GN LM33.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99242614; PubMed=10225960;  
 RA Koch M., Olson P.F., Albus A., Jin W., Hunter D.D., Brunken W.J.,  
 RT Bureson R.E., Champliand M.F.;  
 RT Characterization and expression of the laminin gamma3 chain: a novel,  
 RT non-basement membrane-associated, laminin chain.;  
 RT J. Cell Biol. 145:605-618(1999).  
 CC - FUNCTION: Binding to cells via a high affinity receptor, laminin  
 CC is thought to mediate the attachment, migration, and organization  
 CC of cells into tissues during embryonic development by interacting  
 CC with other extracellular matrix components.  
 CC - SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
 CC different polypeptide chains (alpha, beta, gamma), which are bound  
 CC to each other by disulfide bonds into a cross-shaped molecule  
 CC comprising one long and three short arms with globules at each  
 CC end.  
 CC THE GAMMA-3 CHAIN IS A SUBUNIT OF LAMININ-12.  
 CC - SUBCELLULAR LOCATION: Extracellular.  
 CC - TISSUE SPECIFICITY: Broadly expressed in: skin, heart, lung, and





DB 922 LGSEDOCHPKTCGCTCRPGVGTGACDRGOLFSSGCRACRCSPLGASAGC 977

RESULT 11  
LMB1\_HUMAN STANDARD: PRT: 1786 AA.  
AC P07942;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 15-JUN-2002 (Rel. 08, Last sequence update)  
DE Laminin beta-1 chain precursor (Laminin BI chain).  
CN LMB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368768; PubMed=1975589;  
RA Vuolteenaho R., Chow L.T., Tryggvason K.;  
RT "Structure of the human laminin BI chain gene.";  
RL J. Biol. Chem. 265:15611-15616(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87280097; PubMed=3611077;  
RA Plikarainen T., Eddy R., Fukushima Y., Byers M., Shows T.,  
RT "Human laminin BI chain. A multidomain protein with gene (LAMB1)  
J. Biol. Chem. 262:10454-10462(1987).  
RN [3]  
RP SEQUENCE OF 1276-1709 FROM N.A.  
RX MEDLINE=88021029; PubMed=3661559;  
RA Jaye M., Modi W.S., Rieck G.A., Mudd R., Chiu I.M., O'Brien S.J.,  
RT "Isolation of a cDNA clone for the human laminin-BI chain and its  
gene localization.";  
RL Am. J. Hum. Genet. 41:605-615(1987).  
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin  
is thought to mediate the attachment, migration, and organization  
of cells into tissues during embryonic development by interacting  
with other extracellular matrix components.  
CC -1- SUBUNIT: laminin is a complex glycoprotein, consisting of three  
different polypeptide chains (alpha, beta, gamma), which are bound  
to each other by disulfide bonds into a cross-shaped molecule  
comprising one long and three short arms with globules at each  
end.  
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-  
2 (MEROSIN), AND LAMININ-6 (K-LAMININ).  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
COMPONENT).  
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -1- SIMILARITY: CONTAINS 12.5 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M61951; AAA59486.1; -  
DR EMBL; M58147; AAA59486.1; JOINED.  
DR EMBL; M61917; AAA59486.1; JOINED.  
DR EMBL; M61918; AAA59486.1; JOINED.  
DR EMBL; M61921; AAA59486.1; JOINED.  
DR EMBL; M61922; AAA59486.1; JOINED.

DR EMBL; M61923; AAA59486.1; JOINED.  
DR EMBL; M61924; AAA59486.1; JOINED.  
DR EMBL; M61925; AAA59486.1; JOINED.  
DR EMBL; M61926; AAA59486.1; JOINED.  
DR EMBL; M61927; AAA59486.1; JOINED.  
DR EMBL; M61928; AAA59486.1; JOINED.  
DR EMBL; M61929; AAA59486.1; JOINED.  
DR EMBL; M61930; AAA59486.1; JOINED.  
DR EMBL; M61931; AAA59486.1; JOINED.  
DR EMBL; M61932; AAA59486.1; JOINED.  
DR EMBL; M61933; AAA59486.1; JOINED.  
DR EMBL; M61934; AAA59486.1; JOINED.  
DR EMBL; M61935; AAA59486.1; JOINED.  
DR EMBL; M61936; AAA59486.1; JOINED.  
DR EMBL; M61937; AAA59486.1; JOINED.  
DR EMBL; M61938; AAA59486.1; JOINED.  
DR EMBL; M61939; AAA59486.1; JOINED.  
DR EMBL; M61940; AAA59486.1; JOINED.  
DR EMBL; M61941; AAA59486.1; JOINED.  
DR EMBL; M61942; AAA59486.1; JOINED.  
DR EMBL; M61943; AAA59486.1; JOINED.  
DR EMBL; M61944; AAA59486.1; JOINED.  
DR EMBL; M61945; AAA59486.1; JOINED.  
DR EMBL; M61946; AAA59486.1; JOINED.  
DR EMBL; M61947; AAA59486.1; JOINED.  
DR EMBL; M61948; AAA59486.1; JOINED.  
DR EMBL; M61949; AAA59486.1; JOINED.  
DR EMBL; M61950; AAA59486.1; JOINED.  
DR EMBL; M55370; AAA59485.1; JOINED.  
DR EMBL; M55378; AAA59485.1; JOINED.  
DR EMBL; M55365; AAA59485.1; JOINED.  
DR EMBL; M55371; AAA59485.1; JOINED.  
DR EMBL; M55372; AAA59485.1; JOINED.  
DR EMBL; M55373; AAA59485.1; JOINED.  
DR EMBL; M55374; AAA59485.1; JOINED.  
DR EMBL; M55375; AAA59485.1; JOINED.  
DR EMBL; M55376; AAA59485.1; JOINED.  
DR EMBL; M55344; AAA59485.1; JOINED.  
DR EMBL; M55345; AAA59485.1; JOINED.  
DR EMBL; M55346; AAA59485.1; JOINED.  
DR EMBL; M55347; AAA59485.1; JOINED.  
DR EMBL; M55348; AAA59485.1; JOINED.  
DR EMBL; M55349; AAA59485.1; JOINED.  
DR EMBL; M55350; AAA59485.1; JOINED.  
DR EMBL; M55351; AAA59485.1; JOINED.  
DR EMBL; M55352; AAA59485.1; JOINED.  
DR EMBL; M55353; AAA59485.1; JOINED.  
DR EMBL; M55355; AAA59485.1; JOINED.  
DR EMBL; M55357; AAA59485.1; JOINED.  
DR EMBL; M55358; AAA59485.1; JOINED.  
DR EMBL; M55359; AAA59485.1; JOINED.  
DR EMBL; M55360; AAA59485.1; JOINED.  
DR EMBL; M55361; AAA59485.1; JOINED.  
DR EMBL; M55362; AAA59485.1; JOINED.  
DR EMBL; M55363; AAA59485.1; JOINED.  
DR EMBL; M55364; AAA59485.1; JOINED.  
DR EMBL; M55366; AAA59485.1; JOINED.  
DR EMBL; M55367; AAA59485.1; JOINED.  
DR EMBL; M55368; AAA59485.1; JOINED.  
DR EMBL; M55369; AAA59485.1; JOINED.  
DR EMBL; M61916; AAA59487.1; -  
DR EMBL; M62006; AAA59487.1; -  
DR PIR; S13547; MMBH01.  
DR HSP; P02468; IKT0.  
DR Genew; HGNC:6486; LAMB1.  
DR MIM; 150240; -  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LAMNT.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00053; laminin\_EGF\_13.  
DR Pfam; PF00055; laminin\_Nterm\_1.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR ProDom; PD002082; LaminT; 1.

SMART; SM00180; EGF Lam; 11.  
 DR SMART; SM00180; EGF Lam; 11.  
 DR PROSITE; PS00022; EGF\_1; 9.  
 DR PROSITE; PS01248; EGF\_2; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
 DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
 FT CHAIN 1 21 LAMININ BETA-1 CHAIN.  
 FT DOMAIN 22 1786 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 22 270 LAMININ EGF-LIKE 1.  
 FT DOMAIN 271 334 LAMININ EGF-LIKE 2.  
 FT DOMAIN 335 397 LAMININ EGF-LIKE 3.  
 FT DOMAIN 398 457 LAMININ EGF-LIKE 4.  
 FT DOMAIN 458 509 LAMININ EGF-LIKE 5 (INCOMPLETE).  
 FT DOMAIN 510 540 LAMININ DOMAIN IV.  
 FT DOMAIN 541 771 LAMININ EGF-LIKE 6.  
 FT DOMAIN 773 820 LAMININ EGF-LIKE 7.  
 FT DOMAIN 821 866 LAMININ EGF-LIKE 8.  
 FT DOMAIN 867 916 LAMININ EGF-LIKE 9.  
 FT DOMAIN 917 975 LAMININ EGF-LIKE 10.  
 FT DOMAIN 1027 1083 LAMININ EGF-LIKE 11.  
 FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1179 1397 DOMAIN II.  
 FT DOMAIN 1398 1430 DOMAIN ALPHA.  
 FT DOMAIN 1431 1786 COILED COIL (POTENTIAL).  
 FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).  
 FT DOMAIN 1353 1388 COILED COIL (POTENTIAL).  
 FT DOMAIN 1442 1781 BY SIMILARITY.  
 FT DISULFID 271 280 BY SIMILARITY.  
 FT DISULFID 273 298 BY SIMILARITY.  
 FT DISULFID 300 309 BY SIMILARITY.  
 FT DISULFID 312 332 BY SIMILARITY.  
 FT DISULFID 335 344 BY SIMILARITY.  
 FT DISULFID 337 362 BY SIMILARITY.  
 FT DISULFID 365 374 BY SIMILARITY.  
 FT DISULFID 377 395 BY SIMILARITY.  
 FT DISULFID 398 411 BY SIMILARITY.  
 FT DISULFID 400 426 BY SIMILARITY.  
 FT DISULFID 428 437 BY SIMILARITY.  
 FT DISULFID 440 455 BY SIMILARITY.  
 FT DISULFID 458 472 BY SIMILARITY.  
 FT DISULFID 460 479 BY SIMILARITY.  
 FT DISULFID 481 490 BY SIMILARITY.  
 FT DISULFID 493 507 BY SIMILARITY.  
 FT DISULFID 507 575 BY SIMILARITY.  
 FT DISULFID 773 785 BY SIMILARITY.  
 FT DISULFID 794 803 BY SIMILARITY.  
 FT DISULFID 806 818 BY SIMILARITY.  
 FT DISULFID 821 833 BY SIMILARITY.  
 FT DISULFID 833 840 BY SIMILARITY.  
 FT DISULFID 842 851 BY SIMILARITY.  
 FT DISULFID 854 864 BY SIMILARITY.  
 FT DISULFID 867 876 BY SIMILARITY.  
 FT DISULFID 883 895 BY SIMILARITY.  
 FT DISULFID 886 895 BY SIMILARITY.  
 FT DISULFID 898 914 BY SIMILARITY.  
 FT DISULFID 933 933 BY SIMILARITY.

Query Match 3.2%; Score 174; DB 1; Length 1786;  
 Best Local Similarity 20.9%; Pred. No. 0.00018; Indels 290; Gaps 46;  
 Matches 168; Conservative 69; Mismatches 278;

OY 275 CFPCKPGTY----ADKQSSPCKICPANSYNSKGETSCHQCDPKYSE-----KGSSEC 324  
 DB 374 CECKPFFYQHPRDIRPNCFCRCKTCDDPAGSNEGIC-----DSTYDFSTGLIAGQCRC 428  
 OY 325 --NVRPACTD--KDYFY-----THTACDANG-----ETQIMY----- 352  
 DB 429 KLVNBEHCDVCKEGFYDLSSEDPGCKSCACNPLGTTPGANPCDSTGHCYCKRLVTVGO 488  
 OY 353 -----KWAPKICSEDLBGAVKLPASGVKTHCPC-----NPGFFFTNNSTCO 395

DB 489 HCDQCLPEHNG-----LSNDLDG-----CRPDGDLGALNNSCFAESGQSCSR 532  
 OY 396 PCPYGYSN-----GSDTCRCP 412  
 DB 533 PHMIGROCNVEEGYFATLDHLYAEAEANLPGVSIYEROYIODRIPMTGAGFYRVF 592  
 OY 413 AGTEPAVGEFKWMNTLPMTETVLSGINFEEKMTGWEVAGDHITVYAGASNDNFML 472  
 DB 593 EGAY-----LIEF-FIDNIPYSMEYDIL--IKTEPOLPDHWEK-----VI 629  
 OY 473 TLVVPGEFRPPOS-----VMADPENKEV-----ARITVEFTLGSVNCLELYFVGVN--SRT 521  
 DB 630 TVORGPRIPTSSRCGNITIPDDNOVSLSPGSRYVILPRVC-----FEGCTIVYTRL 682  
 OY 522 NTPVETWKGSKOSQSYIIE-----ENTTISFTV-AFORTTFEHA 561  
 DB 683 ELPOYTSSDSDVESPYTLIDSLVLMPYCKSLDITTVGSGDGVVTSAMETFORVRCLEN 742  
 OY 562 SRKYT-----NDYAK--IYSINVTVMNGVASYCRP-----CALFASDVSS 601  
 DB 743 SRSVVKTPMTDVCNRLITISALHQTGLACECDPOGSLSSVCDPNGGOCRCRPNVGR 802  
 OY 602 CTSCPAGYIYDRSG-----TCH-----SCPMTILKAHOPYVOA-----CYP----- 640  
 DB 803 CNRCAPCTFEGFPGSGCKPCECHLOGSVNAFCNPVT--GCHCFQGVYAROCDCRCLPGHWF 861  
 OY 641 --GSPGKKNKIKHSLYNDCTESRNTPTREYVNSALANVTLLAGGSPFSKLGKLYHN 698  
 DB 862 PSCPPQCCNG--HA--DDC-----DPVT-----GRCLMCDQYTMGN 894  
 OY 699 FTLSLGNCKGKMSVCTDNTDRIPEG-ESG--FESKI-----TAYVQAVIIPPE 747  
 DB 895 CERCIAGYGGPILIGSDHCRPCPCPDGPRGQFARSCYODPVTLQALCVCDDPCYICSR 954  
 OY 748 ----VTGY-----KAGVSSOPVSLADRLIVTDTMLDITGTPAE--LFLHLESLGIPVY 796  
 DB 955 CDDCASTYFGNPSEVGGSCQPCQCHNN-IDTTPPEACDKETGCKLCLVHTGEGHCQFCR 1013  
 OY 797 FFYRSNDYQSSS-----GRSTTRVRCSP-----OKTVPGLLPCTGSDGCDGCGNPF 848  
 DB 1014 FGYYGALMLQDCKRCVCNLTGVQEHGNSDQCDKATGCIQLPNVTVGO--NCDRCAPN-- 1071  
 OY 849 LME--SAAACPLCSVADYHAIYSSC 871  
 DB 1072 TWQLASGTGDCPCNCMAHSFSPSC 1096

RESULT 12  
 LMA2\_MOUSE STANDARD; PRT; 3106 AA.  
 ID LMA2\_MOUSE  
 AC 060675; Q05003; Q64061;  
 DT 01-NOV-1997 (Rel. 35, last sequence update)  
 DT 01-NOV-1997 (Rel. 35, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy chain).  
 GN LMA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_Taxid=10090;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Embryo, and Heart;  
 MEDLINE=95316259; PubMed=7795883;  
 RA Yamada Y.; Utani A.; Sugiyama S.; Doi T.; Polistina C.;  
 RT "Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse."  
 RI Matrix Biol. 14:447-455(1995).  
 RN [2]  
 RP SEQUENCE OF 2162-2279 FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Thymus;  
RA MEDLINE=93346725; PubMed=8345183;  
RA Chang A.C., Wadsworth S., Coligan J.E.;  
RT "Expression of merosin in the thymus and its interaction with  
RT thymocytes";  
RL J. Immunol. 151:1789-1801(1993).  
RN [3]  
RP SEQUENCE OF 64-281 FROM N.A.  
RX MEDLINE=95179178; PubMed=7874173;  
RA Xu H., Wu X.R., Wewer U.M., Engvall E.;  
RT "Murine muscular dystrophy caused by a mutation in the laminin alpha  
RT 2 (Lama2) gene";  
RL Nat. Genet. 8:297-302(1994).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.  
RX MEDLINE=20085745; PubMed=10619025;  
RA Hohenester E., Tisi D., Tait J.F., Timpl R.;  
RT "The crystal structure of a laminin G-like module reveals the  
RT molecular basis of alpha-dystroglycan binding to laminins, perlecan,  
RT and agrin";  
RL Mol. Cell 4:783-792(1999).  
CC -1- FUNCTION: Binding to cells via a high affinity receptor. Laminin  
CC is thought to mediate the attachment, migration, and organization  
CC of cells into tissues during embryonic development by interacting  
CC with other extracellular matrix components.  
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
CC different polypeptide chains (alpha, beta, gamma), which are bound  
CC to each other by disulfide bonds into a cross-shaped molecule  
CC comprising one long and three short arms with globules at each  
CC end.  
CC THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-  
CC 4 (S-MEROSIN).  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
CC COMPONENT).  
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -1- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MURINE MUSCULAR DYSTROPHY  
CC (DY2J).  
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U012147; AAC52165.1; -  
DR EMBL: X69869; CAA49502.1; -  
DR EMBL: S75315; AAB33573.1; -  
DR PDB: 1OU0; 03-DEC-99.  
DR MGD: MGI:99912; Lama2.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001886; LamNT.  
DR InterPro: IPR000034; Laminin\_B.  
DR InterPro: IPR001791; Laminin\_G.  
DR Pfam: PF00052; Laminin\_B\_2.  
DR Pfam: PF00053; Laminin\_EGF\_15.  
DR Pfam: PF00054; Laminin\_G\_5.  
DR Pfam: PF00055; Laminin\_Nterm\_1.  
DR PRINTS: PR00011; EGF\_LAMININ.  
DR PRODOM: PD002082; LamNT\_1.  
DR PRODOM: PD003031; Laminin\_B\_2.  
DR SMART: SM00180; EGF\_Lam\_15.  
DR SMART: SM00001; EGF\_Like\_1.  
DR SMART: SM00281; Lamb\_2.

DR SMART; SM00282; Lamb; 5.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF-1; 11.  
DR PROSITE; PS01186; EGF-2; 3.  
DR PROSITE; PS01246; LAMININ\_TYPE\_EGF; 14.  
DR PROSITE; PS50025; LAM\_G\_DOMAIN; 5.  
DR Glycoprotein; Basement membrane; Cell adhesion; Repeat; Signal; 3D-structure.  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 22  
FT CHAIN 23 3106  
FT DOMAIN 23 282 LAMININ ALPHA-2 CHAIN.  
FT DOMAIN 283 336 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 340 409 LAMININ EGF-LIKE 1.  
FT DOMAIN 410 464 LAMININ EGF-LIKE 2.  
FT DOMAIN 465 513 LAMININ EGF-LIKE 3.  
FT DOMAIN 514 523 LAMININ EGF-LIKE 4.  
FT DOMAIN 524 719 LAMININ EGF-LIKE 5 (N-TERMINAL).  
FT DOMAIN 720 752 LAMININ DOMAIN IV 1 (DOMAIN IV B).  
FT DOMAIN 753 802 LAMININ EGF-LIKE 5 (C-TERMINAL).  
FT DOMAIN 803 860 LAMININ EGF-LIKE 6.  
FT DOMAIN 861 913 LAMININ EGF-LIKE 7.  
FT DOMAIN 914 962 LAMININ EGF-LIKE 8.  
FT DOMAIN 963 1009 LAMININ EGF-LIKE 9.  
FT DOMAIN 1010 1055 LAMININ EGF-LIKE 10.  
FT DOMAIN 1056 1101 LAMININ EGF-LIKE 11.  
FT DOMAIN 1102 1161 LAMININ EGF-LIKE 12.  
FT DOMAIN 1162 1171 LAMININ EGF-LIKE 13.  
FT DOMAIN 1172 1375 LAMININ EGF-LIKE 14 (N-TERMINAL).  
FT DOMAIN 1376 1415 LAMININ DOMAIN IV 2 (DOMAIN IV A).  
FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15.  
FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.  
FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.  
FT DOMAIN 1570 2140 LAMININ II AND I.  
FT DOMAIN 2141 2324 LAMININ G-LIKE 1.  
FT DOMAIN 2326 2517 LAMININ G-LIKE 2.  
FT DOMAIN 2522 2706 LAMININ G-LIKE 3.  
FT DOMAIN 2759 2930 LAMININ G-LIKE 4.  
FT DOMAIN 2929 3106 LAMININ G-LIKE 5.  
FT DOMAIN 1662 1863 COILED COIL (POTENTIAL).  
FT DOMAIN 1923 2146 COILED COIL (POTENTIAL).  
FT DISULFID 285 292 BY SIMILARITY.  
FT DISULFID 305 314 BY SIMILARITY.  
FT DISULFID 317 337 BY SIMILARITY.  
FT DISULFID 340 349 BY SIMILARITY.  
FT DISULFID 377 374 BY SIMILARITY.  
FT DISULFID 389 407 BY SIMILARITY.  
FT DISULFID 410 422 BY SIMILARITY.  
FT DISULFID 412 438 BY SIMILARITY.  
FT DISULFID 440 449 BY SIMILARITY.  
FT DISULFID 452 462 BY SIMILARITY.  
FT DISULFID 465 478 BY SIMILARITY.  
FT DISULFID 467 482 BY SIMILARITY.  
FT DISULFID 484 493 BY SIMILARITY.  
FT DISULFID 496 511 BY SIMILARITY.  
FT DISULFID 511 762 BY SIMILARITY.  
FT DISULFID 753 769 BY SIMILARITY.  
FT DISULFID 772 781 BY SIMILARITY.  
FT DISULFID 784 800 BY SIMILARITY.  
FT DISULFID 803 818 BY SIMILARITY.  
FT DISULFID 805 828 BY SIMILARITY.  
FT DISULFID 831 840 BY SIMILARITY.  
FT DISULFID 843 858 BY SIMILARITY.  
FT DISULFID 861 875 BY SIMILARITY.  
FT DISULFID 863 882 BY SIMILARITY.  
FT DISULFID 885 894 BY SIMILARITY.  
FT DISULFID 897 911 BY SIMILARITY.  
FT DISULFID 914 926 BY SIMILARITY.  
FT DISULFID 916 933 BY SIMILARITY.  
FT DISULFID 935 944 BY SIMILARITY.  
FT DISULFID 947 960 BY SIMILARITY.  
FT DISULFID 963 975 BY SIMILARITY.

```

FT DISULFID 965 961 BY SIMILARITY.
FT DISULFID 983 992 BY SIMILARITY.
FT DISULFID 995 1007 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1012 1026 BY SIMILARITY.
FT DISULFID 1028 1037 BY SIMILARITY.
FT DISULFID 1040 1053 BY SIMILARITY.
FT DISULFID 1056 1068 BY SIMILARITY.
FT DISULFID 1058 1075 BY SIMILARITY.
FT DISULFID 1077 1086 BY SIMILARITY.
FT DISULFID 1089 1099 BY SIMILARITY.
FT DISULFID 1416 1425 BY SIMILARITY.
FT DISULFID 1418 1432 BY SIMILARITY.
FT DISULFID 1435 1444 BY SIMILARITY.
FT DISULFID 1447 1462 BY SIMILARITY.
FT DISULFID 1465 1480 BY SIMILARITY.
FT DISULFID 1467 1490 BY SIMILARITY.
FT DISULFID 1493 1502 BY SIMILARITY.
FT DISULFID 1505 1520 BY SIMILARITY.
FT DISULFID 1523 1535 BY SIMILARITY.
FT DISULFID 1525 1542 BY SIMILARITY.
FT DISULFID 1544 1553 BY SIMILARITY.
FT DISULFID 1556 1567 BY SIMILARITY.
FT DISULFID 1570 1570 INTERCHAIN (PROBABLE).
FT DISULFID 1574 1574 INTERCHAIN (PROBABLE).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match 3.1%; Score 172.5; DB 1; Length 3106;

Best Local Similarity 19.5%; Pred. No. 0.0047; Mismatches 304; Indels 403; Gaps 57;

```

QY 42 TGPLHACKSEY-----HYEY---ACDSTGRMVAVPHTPG----- 77
DB 836 TGRPCRCAGRYGQPSVPGSGCPQCQCNNDLSIPGSDSLSGSCCLCKPCTTGRYCE 895
QY 78 LCTS--LPDPYKGTGEC-----SFS--CN--AGEFLDMKDQSKPCAERY----- 116
DB 896 LCADYGGDAVNMFKNCQPCRCRDJNGSFESECHRTGQC-----ECRPYNGHHCDECKR 949
QY 117 -----SIGTGIRPEMELP---HGFASSLANNELDSDAESTGNGTSSKMYVRGDIYAF 168
DB 950 ETEGLQIGRC-----LPCNCSFESKSF-----CEASGQC-----WCOPG--VAGK 990
QY 169 NTDECTATLMYAVNLKQSGTVNEFYYPDSIIFEEFVQNDQCQPNADDSRMKKTTEKGW 228
DB 991 KCDRCA-----HGTFNQ-----EGGCIACDCHLGNNDP-----KT----- 1023
QY 229 EHSVELNRGNANLYWMTAFSVWTKVPRVLVNRNIAITGVAVTSECFPCPKPTGYADKOG 288
DB 1024 -----QCQICPPMTG-----EKCESECLPMTW-----G 1046
QY 289 SSF---CKLCPANYS-----NKGETSHOCDPKYSEKSSSON-----VRPACTDK 333
DB 1047 HSIYTGKVCNCSTVSGLASQCUNVNTGQCSH---PKFSGMKCSFESRGHMVPICTTLC 1102
QY 334 DYFY-----THACDANGETQOLMYWAKPKICSEDLG--AVKLPASGKVTQCPNCGFF-- 387
DB 1103 DCFLPQGDATTCDL--ETR-----KCSQSOTGQCSKVNVEG--HCDRCRGRKFG 1150
QY 388 --KTNNSTQCPQPYGYSNG--SDCTRCPAETEPANGFEYKMMNLTPTNNETTVLSGINF 443
DB 1151 LDANKPLGCSGSC---YCFGTQSQCSF-----AKGL--TRTWVTL--SDEQTLPLV-- 1194
QY 444 EYKMTGMEVAGDHLYTAAGASDNDFMILLVVPGFRRPQSVADPNNKVAH----- 497
DB 1195 -----DEALQHTTKGIA-----FQKPRIVAKMDVROELHLEPPYWK 1232
QY 498 --TFVEETLCSVNCCLYFMVGVNSRTNPTVETWKG-----KGQSYTYLIEENTTTSFT 550

```

```

DB 1233 LPQCEGKKMAVGGKLVAYIFEARDETFATYKPOVITIRGSPTHARITRMAAPLI 1292
QY 551 WAFQRTTH--EASKKYINDAKI-----YSINV-----NVMGVASYC 588
DB 1293 GOLTTHETEMKEKEKYGDDPRISRTVTRDFDILDYHILIKATYGVNVR--OSRI 1350
QY 589 RCPALASDVSGSSCSPAGYIDR-----DSG--TCHSCPNTLKAHPYV-----V 635
DB 1351 SETSMEVABPQHVLVAGSPPAHLEKDCDPRGYSGLSETCAGPFRLRSEPGRTPTL 1410
QY 636 QACVPC--GPQTKNNKTHSLCYN-----DCTFSRN 663
DB 1411 CTCVPCQCNHSSQDDPETSVCQNCQHNHTAGDFERCALGYGYIRGLPNDQCPACPLI 1470
QY 664 TPTRTFNNFNSALANTVTLVLAGSPSTSGIKYFHFTSLC--GNQGRKMSVCTDNVDL 721
DB 1471 SPNNFS-----PSCVLEGLE--DYRCTACPRGEGGYCERCACAGYTG 1512
QY 722 RPEGESGFSKSTAYVCOAVIIPPEVTGYKAGVSSQPSVLADRLIGVTTMTLDGITS 781
DB 1513 PSSPFGS-----COECRCDP-----YSLFVP--CDRYTGICT----- 1543
QY 782 AELFHLIESLGIPIVIFYKSNVDYQSCSSGSRSTTRVRCSPQKTYPGSLLPCTCSDGTC 841
DB 1544 -----CRGCAAT-----GRKC 1553
QY 842 DGCNHFLEMSAACPLC 859
DB 1554 DCCE--HWHAREGACVFC 1570

```

RESULT 13

```

ID TS11_GIALA STANDARD; PRT; 667 AA.
AC Q03185;
DR 01-JUN-1994 (Rel. 29, Created)
DR 11-JUN-1994 (Rel. 29, Last sequence update)
DR 15-JUN-1998 (Rel. 36, Last annotation update)
DE Major surface trophozoite antigen 11 precursor.
GN TSP11.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX NCBI_TaxId=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate AD-1;
RX MEDLINE=93241215; PubMed=8479449;
RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;
RT "A gene encoding a 69-kilodalton major surface protein of Giardia
  intestinalis trophozoites.";
RL Mol. Biochem. Parasitol. 58:247-258(1993).
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -I- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.
CC -I- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL, M95814; AA02687.1; -.
CC PIR, A48579; A48579.
CC HSR, P02468; IRL0.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR005127; Giardia_VSP.
CC Pfam, PF03302; VSP; 1.
CC SMART, SM00181; BGF; 3.

```





```

FT DOMAIN 398 457 LAMININ EGF-LIKE 3.
FT DOMAIN 509 509 LAMININ EGF-LIKE 4.
FT DOMAIN 510 540 LAMININ EGF-LIKE 5. (INCOMPLETE).
FT DOMAIN 541 772 LAMININ DOMAIN IV.
FT DOMAIN 773 820 LAMININ EGF-LIKE 6.
FT DOMAIN 821 866 LAMININ EGF-LIKE 7.
FT DOMAIN 867 916 LAMININ EGF-LIKE 8.
FT DOMAIN 917 975 LAMININ EGF-LIKE 9.
FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.
FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.
FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
FT DOMAIN 1179 1397 DOMAIN II.
FT DOMAIN 1398 1430 DOMAIN ALPHA.
FT DOMAIN 1431 1786 DOMAIN I.
FT DOMAIN 1448 1778 COILED COIL (POTENTIAL).
FT DISULFID 271 280 COILED COIL (POTENTIAL).
FT DISULFID 273 298 COILED COIL (POTENTIAL).
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 312 332 BY SIMILARITY.
FT DISULFID 335 344 BY SIMILARITY.
FT DISULFID 337 362 BY SIMILARITY.
FT DISULFID 365 374 BY SIMILARITY.
FT DISULFID 377 395 BY SIMILARITY.
FT DISULFID 398 411 BY SIMILARITY.
FT DISULFID 400 426 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 440 455 BY SIMILARITY.
FT DISULFID 458 472 BY SIMILARITY.
FT DISULFID 460 479 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
FT DISULFID 493 507 BY SIMILARITY.
FT DISULFID 507 785 BY SIMILARITY.
FT DISULFID 773 792 BY SIMILARITY.
FT DISULFID 794 803 BY SIMILARITY.
FT DISULFID 806 818 BY SIMILARITY.
FT DISULFID 821 833 BY SIMILARITY.
FT DISULFID 823 840 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 864 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 869 883 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 898 914 BY SIMILARITY.
FT DISULFID 917 933 BY SIMILARITY.
FT DISULFID 919 944 BY SIMILARITY.
FT DISULFID 946 955 BY SIMILARITY.
FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 978 997 BY SIMILARITY.
FT DISULFID 1000 1009 BY SIMILARITY.
FT DISULFID 1012 1025 BY SIMILARITY.
FT DISULFID 1084 1096 BY SIMILARITY.
FT DISULFID 1086 1103 BY SIMILARITY.
FT DISULFID 1105 1114 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1132 1144 BY SIMILARITY.
FT DISULFID 1134 1151 BY SIMILARITY.
FT DISULFID 1153 1162 BY SIMILARITY.
FT DISULFID 1165 1176 BY SIMILARITY.
FT DISULFID 1179 1179 BY SIMILARITY.
FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).
FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).
FT CARBOHYD 120 120 INTERCHAIN (PROBABLE).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1041 1041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1195 1195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1279 1279 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 1343 1343 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1533 1533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1542 1542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1531 1534 SGNA -> MEMB (IN REF. 2).
FT CONFLICT 1749 1749 D -> N (IN REF. 2).
SQ SEQUENCE 1786 AA; 196904 MW; 846671B7BF41A474 CRC64;

Query Match 3.1%; Score 170; DB 1; Length 1786;
Best Local Similarity 20.2%; Pred. No. 0.00035;
Matches 166; Conservative 73; Mismatches 258; Indels 326; Gaps 48;

OY 275 CFPKPGTY-----ADKQSSFCCLC---PANS-----YSN-----KGETSC----- 308
DB 374 CEQCKPFYFQHPERDRIKPNICCEPCTCPAGSENGICDGTIDFVSLIAGQCKRLIVE 433
OY 309 -HQC-----PDKYEKSSSCNV-----RPACTDKDYF-----THPACDA 344
DB 434 GERCDVCKEGFYDLAEDPYGCK-SCACNPIGTIPGNCPCSEGTGYCKRLVTGQRDC- 491
OY 345 NGETQAMTKAKPKICSEDLGAVKLPASGVKTHCPNCPGFFKTNSTCQ---PCPYG 400
DB 492 --OCLPQHWG---LSNDLDG-----CRPCDCLGALNNSCSEDSGCSCCL 532
OY 401 SYNSGSDCT-----RCPAGTEPAVGF-- 421
DB 533 PHMIGRQCNVESEGYFTTLHDHYIEAEANIGPCVVVVEROYIDRIIPSWTGP--GVYR 590
OY 422 -----EYKMWNLPTMETTVLSGINFEXKGMTGWEVAGDHITYAGASDNDEMIITL 474
DB 591 VPBGAYLEF-FLDNPYMEYEIL--IYEPOLPDHMEKA-----VITV 631
OY 475 VVPGFRPQS-----VMATENKEV-----ARTYFELCSYNCGLYMGVY--SRNT 523
DB 632 QPFGKIPASSCGMTVPDDNOVYSLSPGSRVYLPBVC-----FEKGNYTVREL 684
OY 524 PVEWKGSGKOSQSYTYE-----ENTTSFTW-AFORTFHEASR 563
DB 685 PQTASGSDVSPPTFIDSLVLMPCSKSLDFTVVGSGDGEVTNAMETPQRYRLEMSR 744
OY 564 KYT---NDVAK--IYSINVTVMNGVASYCRP-----CALEASDVSSCT 603
DB 745 SVVTPMTDVCNRNIIFFSALIHQGLACEDDQGLSVCDPNCGQCCQCRNVVYGRICN 804
OY 604 SCPAG-----YIYDSDGTC-----SCPNTTLKAHQ 631
DB 805 RCAGTGFEGFGNGCKPCDCHLQGSASAFCDATGQCHCFQGIYARQDCRCLPGY----- 858
OY 632 PYGVQACVPCGPRGPKNNKIHSLCYNDCTFFSNTPRTFTNYNFSALANTVTLAGSPFSK 691
DB 859 -WGFSPQPC-----QCNGHAL--DC-----DVTY--GCCLSCQ 887
OY 692 GLKFFHFFTLSCGNQGRKMSVCTDNYDLRIPEG-ESG--FSKSI-----TAYVCO 740
DB 888 DYTGHNERLQAGYGPRIIGSGDHCRPCRPDRSDGRQFARBCYDDPYTLQACVCD 947
OY 741 AVIIPREYTKKAGVSQVSLADRLIGVTTDMTIDGITSFAELFHELSLGIPTY----- 795
DB 948 -----PGY-----IGSRCDCCASGFFGNPSD--FGSGCQPCQCHNHNIDTTPREACDRPT 994
OY 796 -----IIFYRNDYQSGSSG-----RSTTIYRNCSPQKTYR-----GSLT--LPPTC 836
DB 995 GRCLKCLYHTGDRHQCLQYGYGDALQDCKRCVCNLTGVYKEHCNSDCHCDKATQGC 1054
OY 837 S-----DGTDCGCFHFLME--SAAACPLCSVADYHAIIVSC 871
DB 1055 SCLNIVGNQNDRCARP--TWQLASGTGCGPCNCAHAHSFGFSC 1096

RESULT 15
FUR2_DROME
ID FUR2_DROME STANDARD; PRT; 1680 AA.

```

AC P30432;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).  
 GN FUR2.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92381036; PubMed=1512259;  
 RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,  
 RA Reutrop M., Gattef E.A.F., Leunissen J.A.M., van de Ven W.J.;  
 RT "Cloning and functional expression of Drosophila melanogaster with multiple  
 RT proteolytic processing enzyme of Drosophila melanogaster with multiple  
 RT repeats of a cysteine motif.";  
 RL J. Biol. Chem. 267:17208-17215(1992).  
 CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPROTEASE  
 CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF  
 CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their  
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa  
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,  
 CC complement component C3 and von Willebrand factor from their  
 CC respective precursors.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL: M94375; AAA28551.1; -;  
 DR PIR: A43434; A43434.  
 DR HSP: O99405; 1MPT.  
 DR MEROPS: S08.049; -;  
 DR FLYbase: FBgn0004598; Fur2.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002884; P domain.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR Pfam: PF01483; P; PARTIAL.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR ProDom: PR000717; P domain; 1.  
 DR SMART: SM00181; EGF 1.  
 DR SMART: SM00261; FU; 10.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 KW Hydroxylase; Serine protease; Glycoprotein; Signal; Transmembrane;  
 KW Multigene family; Zymogen; Repeat.  
 FT SIGNAL 1  
 FT PROPEP 1  
 FT CHAIN 320 1680  
 FT ACT\_SITE 418 418  
 FT ACT\_SITE 457 457  
 FT ACT\_SITE 638 638  
 FT DOMAIN 962 1444  
 FT REPEAT 962 1007  
 FT REPEAT 1008 1057  
 FT REPEAT 1058 1104  
 FT REPEAT 1105 1153  
 FT REPEAT 1154 1205  
 FT REPEAT 1206 1254  
 FT REPEAT 1255 1299  
 FT REPEAT 1300 1346

FT REPEAT 1347 1393  
 FT REPEAT 1394 1444  
 FT TRANSMEM 1508 1532  
 FT DOMAIN 1533 1680  
 FT CARBOHYD 109 109  
 FT CARBOHYD 130 130  
 FT CARBOHYD 203 203  
 FT CARBOHYD 443 443  
 FT CARBOHYD 481 481  
 FT CARBOHYD 928 928  
 FT CARBOHYD 1061 1061  
 FT CARBOHYD 1182 1182  
 FT CARBOHYD 1275 1275  
 FT CARBOHYD 1278 1278  
 FT CARBOHYD 1440 1440  
 SQ SEQUENCE 1680 AA; 183599 MW; 0A99CE8770A8E293 CRC64;  
 Query Match 3.1%; Score 169; DB 1; Length 1680;  
 Best Local Similarity 17.2%; Pred. No. 0.00038;  
 Matches 159; Conservative 85; Mismatches 291; Indels 388; Gaps 41;  
 QY 46 LKAKESEHYEXTACDSG-----SRMY-----AVPHPGCTSL 82  
 DB 965 LHSCD-----AECDSGGYGRGPTQCVASCNHRKLDNTCVSRCPSPRQVQVQVQV 1016  
 QY 83 PPVVKTECFSCNAGEFLDMKDOCKPCAEGR--YSLGTGIRFDEMDELPHGFASLSA 139  
 DB 1017 HD-----TCETCAGAG-----PDSCLTCAPAHMLHYIDLAVCLQF-----CPDGYENSR 1060  
 QY 140 NMELEDSAEISGNCSTSKWVRPGDYIAENTDECATYLMAYNKLKSGVNEFYIYDSS 199  
 DB 1061 NR-----TCVPCBPNCASCODHP-----EYCTSC-----DHLVMEHKCYASC 1099  
 QY 200 IIEFFVQNDQO-----PNAEDSRMKKTEKGEFHSVELNCGNVLYVRTA 248  
 DB 1100 PLDTYETENKCAFCHSTGATCNGPTDQCTICRSRYAWO-----NKLISCPRG 1150  
 QY 249 FSWMTKVPKVLVNRNIAITGVAITYSECPCKPGYADKOGSFECKLPANSSYNNKGTSC 308  
 DB 1151 FYADKK-----RLECMPCOGS-----CKTCTSGNV-----C 1176  
 QY 309 HOC-----DPDKYSEKSSSCNVRPACTDKDYFT-----HTACDANGETQLMY 352  
 DB 1177 SECLQWTLNKRDKCIVSGSECS-----ESEFSQVEGOCRPCHASC----- 1219  
 QY 353 KNAKPRICSEDEGANKVLPASGVKTHCP-----CNPGEFTNNNSCOPCPY- 399  
 DB 1220 -----GSCNGPADTSCSPNRLLEDSRCVSGCREGFVAGSLCSPCLHT 1266  
 QY 400 -GSYNSGDCRCPAGTEPAVGFYKWMNTLPTNMEYVLSGINFEYKGMTGMEVAGDH 458  
 DB 1267 CSQCYSRTKSNCSKGLQNG-----ECFTTADGV-YSDRIGCAKCYLSCH- 1313  
 QY 459 YTAAGASNDDEMLTLVVGFRPPOSVMADTENKEVATTFYFETLCSVNCCLYEMVGN 518  
 DB 1314 -TCSGRRNQ-----ECFTTADGV-YSDRIGCAKCYLSCH- 1313  
 QY 519 S-RTNPTVTKWGSKGKOSTYIIIEENTTSFTWAFQRTTFHASKRYNDVAKIYSIV 577  
 DB 1336 ECHRECPGEGYSDRCQCKH-----AGYUIDRDSGTCHSCP 1357  
 QY 578 TVVMNGVASVCPALADVGS-SCTSCP-----AGYUIDRDSGTCHSCP 622  
 DB 1358 -----YCKTC-----NDAGRLACTSCPRHSMIDGICMCLSSQYIDTTSATKTC- 1403  
 QY 623 PVTILKAHPYGVQACVPCPGCTKNNKHSCLYNDG--TFSRNPTRTFNNYFSALAN- 678  
 DB 1404 HDSRSCFGP-GQFSCGCVPLHLDLQNLNSQCVSCQONQTLAEKTS-----AACCC 1455  
 QY 679 -----TYTLAGGPFSTGKGLYFHHFTLSLQ-----NGRKMVSQTDVTLRIPEGRS 728  
 DB 1456 DGETGECKATSTGKRRITVVGSGAVKSSSKHGFSENGNAREV-----LRDLSPLT 1509



